



**Europäisches
Patentamt**

**European
Patent Office**

**Office européen
des brevets**

Bescheinigung

Certificate

Attestation

Die angehefteten Unterlagen stimmen mit der ursprünglich eingereichten Fassung der auf dem nächsten Blatt bezeichneten europäischen Patentanmeldung überein.

The attached documents are exact copies of the European patent application described on the following page, as originally filed.

Les documents fixés à cette attestation sont conformes à la version initialement déposée de la demande de brevet européen spécifiée à la page suivante.

Patentanmeldung Nr. Patent application No. Demande de brevet n°

04090040.9

Der Präsident des Europäischen Patentamts;
Im Auftrag

For the President of the European Patent Office

Le Président de l'Office européen des brevets
p.o.

R C van Dijk



Anmeldung Nr:
Application no.: 04090040.9
Demande no:

Anmeldetag:
Date of filing: 10.02.04
Date de dépôt:

Anmelder/Applicant(s)/Demandeur(s):

Epigenomics AG
Kastanienallee 24
10435 Berlin
ALLEMAGNE

Bezeichnung der Erfindung/Title of the invention/Titre de l'invention:
(Falls die Bezeichnung der Erfindung nicht angegeben ist, siehe Beschreibung.
If no title is shown please refer to the description.
Si aucun titre n'est indiqué se référer à la description.)

Methods and nucleic acids for the analysis of CpG dinucleotide methylation status
associated with the development of prostate cancer

In Anspruch genommene Priorität(en) / Priority(ies) claimed /Priorité(s)
revendiquée(s)

Staat/Tag/Aktenzeichen/State/Date/File no./Pays/Date/Numéro de dépôt:

EP/01.12.03/EP 03090414

Internationale Patentklassifikation/International Patent Classification/
Classification internationale des brevets:

C12Q1/68

Am Anmeldetag benannte Vertragstaaten/Contracting states designated at date of
filing/Etats contractants désignées lors du dépôt:

AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HU IE IT LU MC NL
PT RO SE SI SK TR LI

METHODS AND NUCLEIC ACIDS FOR THE ANALYSIS OF CpG DINUCLEOTIDE METHYLATION STATUS ASSOCIATED WITH THE DEVELOPMENT OF PROSTATE CANCER.

FIELD OF THE INVENTION

The present invention relates to human DNA sequences that exhibit altered methylation patterns (hypermethylation or hypomethylation) in cancer patients. Particular embodiments of the invention provide highly accurate methods for detection and differentiation of prostate carcinomas.

BACKGROUND

Correlation of aberrant DNA methylation with cancer. Aberrant DNA methylation within CpG 'islands' is characterized by hyper- or hypomethylation of CpG dinucleotide sequences leading to abrogation or overexpression of a broad spectrum of genes, and is among the earliest and most common alterations found in, and correlated with human malignancies. Additionally, abnormal methylation has been shown to occur in CpG-rich regulatory elements in intronic and coding parts of genes for certain tumors. In colon cancer, aberrant DNA methylation constitutes one of the most prominent alterations and inactivates many tumor suppressor genes including, *inter alia*, p14ARF, p16INK4a, THBS1, MINT2, and MINT31 and DNA mismatch repair genes such as hMLH1.

Aside from the specific hypermethylation of tumor suppressor genes, an overall hypomethylation of DNA can be observed in tumor cells. This decrease in global methylation can be detected early, far before the development of frank tumor formation. A correlation between hypomethylation and increased gene expression has been determined for many oncogenes.

Prostate cancer. The prostate is a male sex accessory gland, comprising about 30 to 50 branched glands. It is surrounded by a fibroelastic capsule that separates the gland into discrete lobes. The central zone of the organ is composed of pseudo stratified epithelium, the peripheral zone comprises the bulk of the organ and the two tissue types are separated by a transitional zone.

Benign prostate hypertrophy is present in about 50% of men aged 50 or above, and in 95% of men aged 75 or above. Prostate

cancer is a significant health care problem in Western countries with an incidence of 180 per 100,000 in the United States in 1999 (*Cancer J. Clin.*, 49:8, 1999).

Diagnosis and prognosis of prostate cancer; deficiencies of prior art approaches. Different screening strategies have been employed with at least some degree of success to improve early detection of prostate cancer, including determination of levels of prostate specific antigen ("PSA") and digital rectal examination. If a prostate carcinoma is suspected in a patient, diagnosis of cancer is confirmed or excluded by the histological and cytological analysis of biopsy samples for features associated with malignant transformation. The zone of origin of a prostatic cell proliferative disorder is currently determined by the 'PSA density.' PSA density is determined by dividing the weight of the prostate (as estimated by transrectal ultrasound) by the prostate specific antigen levels of the patient. Levels of over 15% percent are considered as indicative of prostate cancer and grounds for a biopsy. The biopsy, in turn, is used for histological and cytological analysis to determine the zone of origin.

However, using routine histological examination, it is often difficult to distinguish benign hyperplasia of the prostate from early stages of prostate carcinoma, even if an adequate biopsy is obtained (McNeal J. E. et al., *Hum. Pathol.* 2001, 32:441-6). Furthermore, small or otherwise insufficient biopsy samples often impede the analysis.

Molecular markers would offer the advantage that they could be used to efficiently analyze even very small tissue samples, and samples whose tissue architecture has not been maintained. Within the last decade, numerous genes have been studied with respect to differential expression among benign hyperplasia of the prostate and different grades of prostate cancer.

However, no single marker has as yet been shown to be sufficient for the diagnosis of prostate tumors in a clinical setting..

Alternatively, high-dimensional mRNA based approaches may, in particular instances, provide a means to distinguish between different tumor types and benign and malignant lesions. However, application of such approaches as a routine diagnostic

tool in a clinical environment is impeded and substantially limited by the extreme instability of mRNA, the rapidly occurring expression changes following certain triggers (e.g., sample collection), and, most importantly, by the large amount of mRNA needed for analysis which often cannot be obtained from a routine biopsy (see, e.g., Lipshutz, R. J. et al., *Nature Genetics* 21:20-24, 1999; Bowtell, D. D. L. *Nature Genetics Suppl.* 21:25-32, 1999).

The GSTP1 gene. The core promoter region of the Glutathione S-Transferase P gene (GSTP1; accession no. NM_000852) has been shown to be hypermethylated in prostate tumor tissue. The glutathione S-transferase pi enzyme is involved in the detoxification of electrophilic carcinogens, and impaired or decreased levels of enzymatic activity (GSTPi impairment) have been associated with the development of neoplasms, particularly in the prostate. Mechanisms of GSTPi impairment include mutation (the GSTP*B allele has been associated with a higher risk of cancer) and methylation.

Prior art GSTP1 studies. Lee et al., in United States Patent No 5,552,277, disclosed that the expression of the glutathione-S-transferase (GST) Pi gene was downregulated in a significant proportion of prostate carcinomas. Moreover, by means of restriction enzyme analysis they were able to show that the promoter region of the of the GSTPi gene was upmethyalted (hypermethylated) in prostate carcinomas as opposed to normal prostate and leukocyte tissue. However, due to the limited and imprecise nature of the analysis technique used (HpaIII digestion, followed by Southern blotting) the exact number and position of the methylated CG dinucleotides were not characterized.

Douglas et al. (WO9955905) used a method comprising bisulfite treatment, followed by methylation specific PCR to show that prostate carcinoma-specific GSTPi hypermethylation was localized to the core promoter regions, and localized a number of CpG positions that had not been characterised by Lee et al.

Herman and Baylin (United States Patent No. 6,017,704) describe the use of methylation specific primers for methylation analysis, and describe a particular primer pair

suitable for the analysis of the corresponding methylated GSTPi promoter sequence.

However, with respect to the use of GSTPi markers, the prior art is limited with respect to the number of GSTPi promoter CpG sequences that have been characterized for differential methylation status. Moreover, there are no disclosures, suggestions or teachings in the prior art of how such markers could be used to distinguish among benign hyperplasia of the prostate and different grades of prostate cancer.

Aberrant genetic methylation has also been observed in several other genes including AR, p16 (CDKN2a/INK4a), CD44, CDH1. Genome wide hypomethylation for example of the LINE-1 repetitive element has also been associated with tumor progression (Santourlidis S ,Florl A ,Ackermann R ,Wirtz HC ,Schulz WA 'High frequency of alterations in DNA methylation in adenocarcinoma of the prostate.' Prostate 1999 May 15;39(3):166-74).

However, use of these genes as alternative or supplemental diagnostic, or otherwise clinically useful markers in a commercial setting has not been enabled. The application of differentially methylated genes to clinically utilizable platforms requires much further investigation into the sensitivity and specificity of the genes. For example, in the case of the gene CD44, a known metastasis suppressor, downregulation was associated with hypermethylation. However the use of this gene as a commercially available marker was not enabled as it was also methylated in normal tissues. See Vis AN Oomen M Schroder FH van der Kwast TH 'Feasibility of assessment of promoter methylation of the CD44 gene in serum of prostate cancer patients.' Mol Urol. 2001 Winter;5(4):199-203.

Development of medical tests. Two key evaluative measures of any medical screening or diagnostic test are its sensitivity and specificity, which measure how well the test performs to accurately detect all affected individuals without exception, and without falsely including individuals who do not have the target disease (predictive value). Historically, many diagnostic tests have been criticized due to poor sensitivity and specificity.

A true positive (TP) result is where the test is positive and the condition is present. A false positive (FP) result is where the test is positive but the condition is not present. A true negative (TN) result is where the test is negative and the condition is not present. A false negative (FN) result is where the test is negative but the condition is not present.

Sensitivity = $TP / (TP + FN)$

Specificity = $TN / (FP + TN)$

Predictive value = $TP / (TP + FP)$

Sensitivity is a measure of a test's ability to correctly detect the target disease in an individual being tested. A test having poor sensitivity produces a high rate of false negatives, i.e., individuals who have the disease but are falsely identified as being free of that particular disease. The potential danger of a false negative is that the diseased individual will remain undiagnosed and untreated for some period of time, during which the disease may progress to a later stage wherein treatments, if any, may be less effective. An example of a test that has low sensitivity is a protein-based blood test for HIV. This type of test exhibits poor sensitivity because it fails to detect the presence of the virus until the disease is well established and the virus has invaded the bloodstream in substantial numbers. In contrast, an example of a test that has high sensitivity is viral-load detection using the polymerase chain reaction (PCR). High sensitivity is achieved because this type of test can detect very small quantities of the virus. High sensitivity is particularly important when the consequences of missing a diagnosis are high.

Specificity, on the other hand, is a measure of a test's ability to identify accurately patients who are free of the disease state. A test having poor specificity produces a high rate of false positives, i.e., individuals who are falsely identified as having the disease. A drawback of false positives is that they force patients to undergo unnecessary medical procedures treatments with their attendant risks, emotional and

financial stresses, and which could have adverse effects on the patient's health. A feature of diseases which makes it difficult to develop diagnostic tests with high specificity is that disease mechanisms, particularly in cancer, often involve a plurality of genes and proteins. Additionally, certain proteins may be elevated for reasons unrelated to a disease state. An example of a test that has high specificity is a gene-based test that can detect a p53 mutation. Specificity is important when the cost or risk associated with further diagnostic procedures or further medical intervention are very high.

The PSA blood test has a sensitivity of 73%, specificity of 60% and predictive value of 31.5%. PSA sensitivity and specificity can be improved but involve tradeoffs. PSA sensitivity can be improved by adjusting the "normal" PSA level to a lower value for younger men or by following serum PSA values in an individual patient over time (PSA velocity). Both methods will increase the number of cancers detected, but they also increase the number of men undergoing biopsy. Conversely, specificity can be improved by using higher "normal" PSA levels for older men, by using the free-to-total PSA ratio, or by adjusting the normal value according to the size of the prostate. These three methods decrease the number of unnecessary biopsies, but they increase the risk that some cancers will be missed.

It can therefore be seen that there exists a need for a means of prostate cancer diagnosis with improved sensitivity, specificity and/or predictive value.

Sensitivity and specificity of quantitative methylation-specific polymerase chain reaction (QMSP) assay alone (without histological analysis) in prostate cancer analysis of needle biopsies has ranged from 30% sensitivity and 100% specificity to 89% sensitivity and 64% specificity (Harden et. al. J Natl Cancer Inst 2003; 95: 1634-1637). However the predictive value of said technique as a clinical screening tool was not analysed. Furthermore, genetic testing of serum and bodily fluids such as urine and saliva would reduce the need for biopsies to detect cancer and would thus be the most effective screening or monitoring tool. However the development of such tests requires an extremely high degree of sensitivity

and specificity. Analysis of GSTP1 gene hypermethylation (Cairns P, Esteller M, Herman JG, Schoenberg M, Jeronimo C, Sanchez-Cespedes M, et al. Molecular detection of prostate cancer in urine by GSTP1 hypermethylation. Clin Cancer Res 2001;7:2727-30.) in urine sediment of prostate cancer patients showed that only 6 out of 22 individuals with elevated methylation levels in biopsied tumors showed corresponding hypermethylation in urine samples.

Multifactorial approach. Cancer diagnostics has traditionally relied upon the detection of single molecular markers (e.g. gene mutations, elevated PSA levels). Unfortunately, cancer is a disease state in which single markers have typically failed to detect or differentiate many forms of the disease. Thus, assays that recognize only a single marker have been shown to be of limited predictive value. A fundamental aspect of this invention is that methylation based cancer diagnostics and the screening, diagnosis, and therapeutic monitoring of such diseases will provide significant improvements over the state-of-the-art that uses single marker analyses by the use of a selection of multiple markers. The multiplexed analytical approach is particularly well suited for cancer diagnostics since cancer is not a simple disease, this multi-factorial "panel" approach is consistent with the heterogeneous nature of cancer, both cytologically and clinically.

Key to the successful implementation of a panel approach to methylation based diagnostic tests is the design and development of optimized panels of markers that can characterize and distinguish disease states. This patent application describes an efficient and unique panel of genes the methylation analysis of one or a combination of the members of the panel enabling the detection of cell proliferative disorders of the prostate with a particularly high sensitivity, specificity and/or predictive value.

Pronounced need in the art. Therefore, in view of the incidence of prostate hyperplasia (50% of men aged 50 or above, and 95% of men aged 75 or above) and prostate cancer (180 per 100,000), there is a substantial need in the art for the development of molecular markers that could be used to

effectively distinguish among benign hyperplasia of the prostate and different grades of prostate cancer. Additionally, there is a pronounced need in the art for the development of molecular markers that could be used to provide sensitive, accurate and non-invasive methods (as opposed to, e.g., biopsy and transrectal ultrasound) for the diagnosis, prognosis and treatment of prostate cell proliferative disorders.

SUMMARY OF THE INVENTION

The disclosed invention provides a means for detection of or differentiation between prostate cell proliferative disorders by analysis of a gene panel, with a sensitivity and specificity suitable for use in a body fluid or serum assay. The present invention provides novel methods for detecting or distinguishing between prostate cell proliferative disorders with a sensitivity of greater than 30% and a specificity of greater than 65%. Said method is most preferably utilised for detecting or detecting and distinguishing between prostate cell proliferative disorders. The invention provides a method for the analysis of biological samples for features associated with the development of prostate cell proliferative disorders, the method characterised in that at least one nucleic acid, or a fragment thereof, from the group consisting of SEQ ID NO: 1 to SEQ ID NO: 20 is/are contacted with a reagent or series of reagents capable of distinguishing between methylated and non methylated CpG dinucleotides within the genomic sequence, or sequences of interest.

The present invention provides a method for ascertaining genetic and/or epigenetic parameters of genomic DNA. The method has utility for the improved diagnosis, treatment and monitoring of prostate cell proliferative disorders, more specifically by enabling the improved identification of and differentiation between subclasses of said disorder and the genetic predisposition to said disorders. The invention presents improvements over the state of the art in that it enables a more specific and sensitive classification of prostate cell proliferative disorders than that achieved by

currently used tests thereby allowing for improved and informed treatment of patients.

Preferably, the source of the test sample is selected from the group consisting of cells or cell lines, histological slides, biopsies, paraffin-embedded tissue, bodily fluids, ejaculate, urine, blood, and combinations thereof. Preferably, the source is biopsies, bodily fluids, ejaculate, urine, or blood.

Specifically, the present invention provides a method for detecting prostate cell proliferative disorders with a sensitivity of greater than 30% and a specificity of greater than 65%, comprising: obtaining a biological sample comprising genomic nucleic acid(s); contacting the nucleic acid(s), or a fragment thereof, with one reagent or a plurality of reagents sufficient for distinguishing between methylated and non methylated CpG dinucleotide sequences within a target sequence of the subject nucleic acid, wherein the target sequence comprises, or hybridizes under stringent conditions to, a sequence comprising at least 16 contiguous nucleotides of SEQ ID NO: 1 to 20, said contiguous nucleotides comprising at least one CpG dinucleotide sequence; and determining, based at least in part on said distinguishing, the methylation state of at least one target CpG dinucleotide sequence, or an average, or a value reflecting an average methylation state of a plurality of target CpG dinucleotide sequences. Preferably, distinguishing between methylated and non methylated CpG dinucleotide sequences within the target sequence comprises methylation state-dependent conversion or non-conversion of at least one such CpG dinucleotide sequence to the corresponding converted or non-converted dinucleotide sequence within a sequence selected from the group consisting of SEQ ID NO: 5 to SEQ ID NO: 20, and contiguous regions thereof corresponding to the target sequence.

Additional embodiments provide a method for the detection of prostate cell proliferative disorders with a sensitivity of greater than 30% and a specificity of greater than 65%, comprising: obtaining a biological sample having subject genomic DNA; extracting the genomic DNA; treating the genomic DNA, or a fragment thereof, with one or more reagents to convert 5-position unmethylated cytosine bases to uracil or to another base that is detectably dissimilar to cytosine in terms

of hybridization properties; contacting the treated genomic DNA, or the treated fragment thereof, with an amplification enzyme and at least two primers comprising, in each case a contiguous sequence at least 9 nucleotides in length that is complementary to, or hybridizes under moderately stringent or stringent conditions to a sequence selected from the group consisting SEQ ID NO: 5 to SEQ ID NO: 20, and complements thereof, wherein the treated DNA or the fragment thereof is either amplified to produce an amplificate, or is not amplified; and determining, based on a presence or absence of, or on a property of said amplificate, the methylation state of at least one CpG dinucleotide sequence selected from the group consisting of SEQ ID NO: 1 to SEQ ID NO: 4, or an average, or a value reflecting an average methylation state of a plurality of CpG dinucleotide sequences thereof. Preferably, at least one such hybridizing nucleic acid molecule or peptide nucleic acid molecule is bound to a solid phase. Preferably, determining comprises use of at least two methods selected from the group consisting of: hybridizing at least one nucleic acid molecule comprising a contiguous sequence at least 9 nucleotides in length that is complementary to, or hybridizes under moderately stringent or stringent conditions to a sequence selected from the group consisting of SEQ ID NO: 5 to SEQ ID NO: 20, and complements thereof; hybridizing at least one nucleic acid molecule, bound to a solid phase, comprising a contiguous sequence at least 9 nucleotides in length that is complementary to, or hybridizes under moderately stringent or stringent conditions to a sequence selected from the group consisting of SEQ ID NO: 5 to SEQ ID NO: 20, and complements thereof; hybridizing at least one nucleic acid molecule comprising a contiguous sequence at least 9 nucleotides in length that is complementary to, or hybridizes under moderately stringent or stringent conditions to a sequence selected from the group consisting of SEQ ID NO: 5 to SEQ ID NO: 20, and complements thereof, and extending at least one such hybridized nucleic acid molecule by at least one nucleotide base; and sequencing of the amplificate.

14 14 Additional embodiments provide novel genomic and chemically modified nucleic acid sequences, as well as oligonucleotides and/or PNA-oligomers for analysis of cytosine

methylation patterns within sequences from the group consisting of SEQ ID NO: 1 to SEQ ID NO: 4 .

BRIEF DESCRIPTION OF THE DRAWINGS

DETAILED DESCRIPTION OF THE INVENTION

Definitions:

The term "Observed/Expected Ratio" ("O/E Ratio") refers to the frequency of CpG dinucleotides within a particular DNA sequence, and corresponds to the $[\text{number of CpG sites} / (\text{number of C bases} \times \text{number of G bases})] \times \text{band length}$ for each fragment.

The term "CpG island" refers to a contiguous region of genomic DNA that satisfies the criteria of (1) having a frequency of CpG dinucleotides corresponding to an "Observed/Expected Ratio" >0.6 , and (2) having a "GC Content" >0.5 . CpG islands are typically, but not always, between about 0.2 to about 1 kb in length.

The term "methylation state" or "methylation status" refers to the presence or absence of 5-methylcytosine ("5-mCyt") at one or a plurality of CpG dinucleotides within a DNA sequence. Methylation states at one or more particular palindromic CpG methylation sites (each having two CpG CpG dinucleotide sequences) within a DNA sequence include "unmethylated," "fully-methylated" and "hemi-methylated."

The term "hemi-methylation" or "hemimethylation" refers to the methylation state of a palindromic CpG methylation site, where only a single cytosine in one of the two CpG dinucleotide sequences of the palindromic CpG methylation site is methylated (e.g., 5'-CC^MGG-3' (top strand): 3'-GGCC-5' (bottom strand)).

The term "hypermethylation" refers to the average methylation state corresponding to an *increased* presence of 5-mCyt at one or a plurality of CpG dinucleotides within a DNA sequence of a test DNA sample, relative to the amount of 5-mCyt found at corresponding CpG dinucleotides within a normal control DNA sample.

The term "hypomethylation" refers to the average methylation state corresponding to a *decreased* presence of 5-mCyt at one or a plurality of CpG dinucleotides within a DNA sequence of a

test DNA sample, relative to the amount of 5-mCyt found at corresponding CpG dinucleotides within a normal control DNA sample.

The term "microarray" refers broadly to both "DNA microarrays," and "DNA chip(s)," as recognized in the art, encompasses all art-recognized solid supports, and encompasses all methods for affixing nucleic acid molecules thereto or synthesis of nucleic acids thereon.

"Genetic parameters" are mutations and polymorphisms of genes and sequences further required for their regulation. To be designated as mutations are, in particular, insertions, deletions, point mutations, inversions and polymorphisms and, particularly preferred, SNPs (single nucleotide polymorphisms).

"Epigenetic parameters" are, in particular, cytosine methylations. Further epigenetic parameters include, for example, the acetylation of histones which, however, cannot be directly analyzed using the described method but which, in turn, correlate with the DNA methylation.

The term "bisulfite reagent" refers to a reagent comprising bisulfite, disulfite, hydrogen sulfite or combinations thereof, useful as disclosed herein to distinguish between methylated and unmethylated CpG dinucleotide sequences.

The term "Methylation assay" refers to any assay for determining the methylation state of one or more CpG dinucleotide sequences within a sequence of DNA.

The term "MS.AP-PCR" (Methylation-Sensitive Arbitrarily-Primed Polymerase Chain Reaction) refers to the art-recognized technology that allows for a global scan of the genome using CG-rich primers to focus on the regions most likely to contain CpG dinucleotides, and described by Gonzalgo et al., *Cancer Research* 57:594-599, 1997.

The term "MethyLight™" refers to the art-recognized fluorescence-based real-time PCR technique described by Eads et al., *Cancer Res.* 59:2302-2306, 1999.

The term "HeavyMethyl™" assay, in the embodiment thereof implemented herein, refers to a HeavyMethyl™ refer to the use of methylation specific *blocking* probes covering CpG positions between the amplification primers.

The term "Ms-SNuPE" (Methylation-sensitive Single Nucleotide Primer Extension) refers to the art-recognized assay described by Gonzalgo & Jones, *Nucleic Acids Res.* 25:2529-2531, 1997.

The term "MSP" (Methylation-specific PCR) refers to the art-recognized methylation assay described by Herman et al. *Proc. Natl. Acad. Sci. USA* 93:9821-9826, 1996, and by US Patent No. 5,786,146.

The term "COBRA" (Combined Bisulfite Restriction Analysis) refers to the art-recognized methylation assay described by Xiong & Laird, *Nucleic Acids Res.* 25:2532-2534, 1997.

The term "MCA" (Methylated CpG Island Amplification) refers to the methylation assay described by Toyota et al., *Cancer Res.* 59:2307-12, 1999, and in WO 00/26401A1.

The term "hybridization" is to be understood as a bond of an oligonucleotide to a complementary sequence along the lines of the Watson-Crick base pairings in the sample DNA, forming a duplex structure.

"Stringent hybridization conditions," as defined herein, involve hybridizing at 68°C in 5x SSC/5x Denhardt's solution/1.0% SDS, and washing in 0.2x SSC/0.1% SDS at room temperature, or involve the art-recognized equivalent thereof (e.g., conditions in which a hybridization is carried out at 60°C in 2.5 x SSC buffer, followed by several washing steps at 37°C in a low buffer concentration, and remains stable). Moderately stringent conditions, as defined herein, involve including washing in 3x SSC at 42°C, or the art-recognized equivalent thereof. The parameters of salt concentration and temperature can be varied to achieve the optimal level of identity between the probe and the target nucleic acid. Guidance regarding such conditions is available in the art, for example, by Sambrook et al., 1989, *Molecular Cloning, A Laboratory Manual*, Cold Spring Harbor Press, N.Y.; and Ausubel et al. (eds.), 1995, *Current Protocols in Molecular Biology*, (John Wiley & Sons, N.Y.) at Unit 2.10.

The terms 'sensitivity' and 'specificity' refer to values calculated with reference to a sample set of male patients with an average age of 65 and a mixed ethnic range including caucasian and african american.

Overview:

The present invention provides for molecular genetic markers that have novel utility for the analysis of methylation patterns associated with the development of prostate cell proliferative disorders with a sensitivity of greater than 30% and a specificity of greater than 65%. Said markers may be used for detecting or distinguishing between prostate cell proliferative disorders, thereby providing improved means for the classification and treatment of said disorders. The markers according to the present invention are analysed in the form of a 'panel' wherein the methylation of one or more genetic sequences of the genes

Bisulfite modification of DNA is an art-recognized tool used to assess CpG methylation status. 5-methylcytosine is the most frequent covalent base modification in the DNA of eukaryotic cells. It plays a role, for example, in the regulation of the transcription, in genetic imprinting, and in tumorigenesis. Therefore, the identification of 5-methylcytosine as a component of genetic information is of considerable interest. However, 5-methylcytosine positions cannot be identified by sequencing, because 5-methylcytosine has the same base pairing behavior as cytosine. Moreover, the epigenetic information carried by 5-methylcytosine is completely lost during, e.g., PCR amplification.

The most frequently used method for analyzing DNA for the presence of 5-methylcytosine is based upon the specific reaction of bisulfite with cytosine whereby, upon subsequent alkaline hydrolysis, cytosine is converted to uracil which corresponds to thymine in its base pairing behavior. Significantly, however, 5-methylcytosine remains unmodified under these conditions. Consequently, the original DNA is converted in such a manner that methylcytosine, which originally could not be distinguished from cytosine by its hybridization behavior, can now be detected as the only remaining cytosine using standard, art-recognized molecular biological techniques, for example, by amplification and hybridization, or by sequencing. All of these techniques are based on differential base pairing properties, which can now be fully exploited.

The prior art, in terms of sensitivity, is defined by a method comprising enclosing the DNA to be analyzed in an agarose matrix, thereby preventing the diffusion and renaturation of the DNA (bisulfite only reacts with single-stranded DNA), and replacing all precipitation and purification steps with fast dialysis (Olek A, et al., A modified and improved method for bisulfite based cytosine methylation analysis, *Nucleic Acids Res.* 24:5064-6, 1996). It is thus possible to analyze individual cells for methylation status, illustrating the utility and sensitivity of the method. An overview of art-recognized methods for detecting 5-methylcytosine is provided by Rein, T., et al., *Nucleic Acids Res.*, 26:2255, 1998.

The bisulfite technique, barring few exceptions (e.g., Zeschnigk M, et al., *Eur J Hum Genet.* 5:94-98, 1997), is currently only used in research. In all instances, short, specific fragments of a known gene are amplified subsequent to a bisulfite treatment, and either completely sequenced (Olek & Walter, *Nat Genet.* 1997 17:275-6, 1997), subjected to one or more primer extension reactions (Gonzalzo & Jones, *Nucleic Acids Res.*, 25:2529-31, 1997; WO 95/00669; U.S. Patent No. 6,251,594) to analyze individual cytosine positions, or treated by enzymatic digestion (Xiong & Laird, *Nucleic Acids Res.*, 25:2532-4, 1997). Detection by hybridization has also been described in the art (Olek et al., WO 99/28498). Additionally, use of the bisulfite technique for methylation detection with respect to individual genes has been described (Grigg & Clark, *Bioessays*, 16:431-6, 1994; Zeschnigk M, et al., *Hum Mol Genet.*, 6:387-95, 1997; Feil R, et al., *Nucleic Acids Res.*, 22:695-, 1994; Martin V, et al., *Gene*, 157:261-4, 1995; WO 9746705 and WO 9515373).

The present invention provides for the use of the bisulfite technique, in combination with one or more methylation assays, for determination of the methylation status of CpG dinucleotide sequences within sequences from the group consisting of SEQ ID NO: 1 to SEQ ID NO: 4. According to the present invention, determination of the methylation status of CpG dinucleotide sequences within sequences from the group consisting of SEQ ID NO: 1 to SEQ ID NO: 4 has diagnostic and prognostic utility.

Methylation Assay Procedures. Various methylation assay procedures are known in the art, and can be used in conjunction with the present invention. These assays allow for determination of the methylation state of one or a plurality of CpG dinucleotides (e.g., CpG islands) within a DNA sequence. Such assays involve, among other techniques, DNA sequencing of bisulfite-treated DNA, PCR (for sequence-specific amplification), Southern blot analysis, and use of methylation-sensitive restriction enzymes.

For example, genomic sequencing has been simplified for analysis of DNA methylation patterns and 5-methylcytosine distribution by using bisulfite treatment (Frommer et al., *Proc. Natl. Acad. Sci. USA* 89:1827-1831, 1992). Additionally, restriction enzyme digestion of PCR products amplified from bisulfite-converted DNA is used, e.g., the method described by Sadri & Hornsby (*Nucl. Acids Res.* 24:5058-5059, 1996), or COBRA (Combined Bisulfite Restriction Analysis) (Xiong & Laird, *Nucleic Acids Res.* 25:2532-2534, 1997).

COBRA. COBRA analysis is a quantitative methylation assay useful for determining DNA methylation levels at specific gene loci in small amounts of genomic DNA (Xiong & Laird, *Nucleic Acids Res.* 25:2532-2534, 1997). Briefly, restriction enzyme digestion is used to reveal methylation-dependent sequence differences in PCR products of sodium bisulfite-treated DNA. Methylation-dependent sequence differences are first introduced into the genomic DNA by standard bisulfite treatment according to the procedure described by Frommer et al. (*Proc. Natl. Acad. Sci. USA* 89:1827-1831, 1992). PCR amplification of the bisulfite converted DNA is then performed using primers specific for the interested CpG islands, followed by restriction endonuclease digestion, gel electrophoresis, and detection using specific, labeled hybridization probes. Methylation levels in the original DNA sample are represented by the relative amounts of digested and undigested PCR product in a linearly quantitative fashion across a wide spectrum of DNA methylation levels. In addition, this technique can be reliably applied to DNA obtained from microdissected paraffin-embedded tissue samples. Typical reagents (e.g., as might be found in a typical COBRA-based kit) for COBRA analysis may include, but are not limited to: PCR primers for specific gene

(or methylation-altered DNA sequence or CpG island); restriction enzyme and appropriate buffer; gene-hybridization oligo; control hybridization oligo; kinase labeling kit for oligo probe; and radioactive nucleotides. Additionally, bisulfite conversion reagents may include: DNA denaturation buffer; sulfonation buffer; DNA recovery reagents or kits (e.g., precipitation, ultrafiltration, affinity column); desulfonation buffer; and DNA recovery components.

Preferably, assays such as "MethyLight•" (a fluorescence-based real-time PCR technique) (Eads et al., *Cancer Res.* 59:2302-2306, 1999), Ms-SNuPE (Methylation-sensitive Single Nucleotide Primer Extension) reactions (Gonzalzo & Jones, *Nucleic Acids Res.* 25:2529-2531, 1997), methylation-specific PCR ("MSP"; Herman et al., *Proc. Natl. Acad. Sci. USA* 93:9821-9826, 1996; US Patent No. 5,786,146), and methylated CpG island amplification ("MCA"; Toyota et al., *Cancer Res.* 59:2307-12, 1999) are used alone or in combination with other of these methods.

MethyLight•. The MethyLight• assay is a high-throughput quantitative methylation assay that utilizes fluorescence-based real-time PCR (TaqMan••) technology that requires no further manipulations after the PCR step (Eads et al., *Cancer Res.* 59:2302-2306, 1999). Briefly, the MethyLight• process begins with a mixed sample of genomic DNA that is converted, in a sodium bisulfite reaction, to a mixed pool of methylation-dependent sequence differences according to standard procedures (the bisulfite process converts unmethylated cytosine residues to uracil). Fluorescence-based PCR is then performed either in an "unbiased" (with primers that do not overlap known CpG methylation sites) PCR reaction, or in a "biased" (with PCR primers that overlap known CpG dinucleotides) reaction. Sequence discrimination can occur either at the level of the amplification process or at the level of the fluorescence detection process, or both.

The MethyLight• assay may be used as a quantitative test for methylation patterns in the genomic DNA sample, wherein sequence discrimination occurs at the level of probe hybridization. In this quantitative version, the PCR reaction provides for unbiased amplification in the presence of a fluorescent probe that overlaps a particular putative

methylation site. An unbiased control for the amount of input DNA is provided by a reaction in which neither the primers, nor the probe overlie any CpG dinucleotides. Alternatively, a qualitative test for genomic methylation is achieved by probing of the biased PCR pool with either control oligonucleotides that do not "cover" known methylation sites (a fluorescence-based version of the "MSP" technique), or with oligonucleotides covering potential methylation sites.

The MethyLight[®] process can be used with a "TaqMan[®]" probe in the amplification process. For example, double-stranded genomic DNA is treated with sodium bisulfite and subjected to one of two sets of PCR reactions using TaqMan[®] probes; e.g., with either biased primers and TaqMan[®] probe, or unbiased primers and TaqMan[®] probe. The TaqMan[®] probe is dual-labeled with fluorescent "reporter" and "quencher" molecules, and is designed to be specific for a relatively high GC content region so that it melts out at about 10°C higher temperature in the PCR cycle than the forward or reverse primers. This allows the TaqMan[®] probe to remain fully hybridized during the PCR annealing/extension step. As the Taq polymerase enzymatically synthesizes a new strand during PCR, it will eventually reach the annealed TaqMan[®] probe. The Taq polymerase 5' to 3' endonuclease activity will then displace the TaqMan[®] probe by digesting it to release the fluorescent reporter molecule for quantitative detection of its now unquenched signal using a real-time fluorescent detection system.

Typical reagents (e.g., as might be found in a typical MethyLight[®]-based kit) for MethyLight[®] analysis may include, but are not limited to: PCR primers for specific gene (or methylation-altered DNA sequence or CpG island); TaqMan[®] probes; optimized PCR buffers and deoxynucleotides; and Taq polymerase.

Ms-SNuPE. The Ms-SNuPE technique is a quantitative method for assessing methylation differences at specific CpG sites based on bisulfite treatment of DNA, followed by single-nucleotide primer extension (Gonzalzo & Jones, *Nucleic Acids Res.* 25:2529-2531, 1997). Briefly, genomic DNA is reacted with sodium bisulfite to convert unmethylated cytosine to uracil while leaving 5-methylcytosine unchanged. Amplification of the desired target sequence is then performed using PCR primers

specific for bisulfite-converted DNA, and the resulting product is isolated and used as a template for methylation analysis at the CpG site(s) of interest. Small amounts of DNA can be analyzed (e.g., microdissected pathology sections), and it avoids utilization of restriction enzymes for determining the methylation status at CpG sites.

Typical reagents (e.g., as might be found in a typical Ms-SNuPE-based kit) for Ms-SNuPE analysis may include, but are not limited to: PCR primers for specific gene (or methylation-altered DNA sequence or CpG island); optimized PCR buffers and deoxynucleotides; gel extraction kit; positive control primers; Ms-SNuPE primers for specific gene; reaction buffer (for the Ms-SNuPE reaction); and radioactive nucleotides. Additionally, bisulfite conversion reagents may include: DNA denaturation buffer; sulfonation buffer; DNA recovery reagents or kit (e.g., precipitation, ultrafiltration, affinity column); desulfonation buffer; and DNA recovery components.

MSP. MSP (methylation-specific PCR) allows for assessing the methylation status of virtually any group of CpG sites within a CpG island, independent of the use of methylation-sensitive restriction enzymes (Herman et al. *Proc. Natl. Acad. Sci. USA* 93:9821-9826, 1996; US Patent No. 5,786,146). Briefly, DNA is modified by sodium bisulfite converting all unmethylated, but not methylated cytosines to uracil, and subsequently amplified with primers specific for methylated versus unmethylated DNA. MSP requires only small quantities of DNA, is sensitive to 0.1% methylated alleles of a given CpG island locus, and can be performed on DNA extracted from paraffin-embedded samples. Typical reagents (e.g., as might be found in a typical MSP-based kit) for MSP analysis may include, but are not limited to: methylated and unmethylated PCR primers for specific gene (or methylation-altered DNA sequence or CpG island), optimized PCR buffers and deoxynucleotides, and specific probes.

MCA. The MCA technique is a method that can be used to screen for altered methylation patterns in genomic DNA, and to isolate specific sequences associated with these changes (Toyota et al., *Cancer Res.* 59:2307-12, 1999). Briefly, restriction enzymes with different sensitivities to cytosine methylation in their recognition sites are used to digest

genomic DNAs from primary tumors, cell lines, and normal tissues prior to arbitrarily primed PCR amplification. Fragments that show differential methylation are cloned and sequenced after resolving the PCR products on high-resolution polyacrylamide gels. The cloned fragments are then used as probes for Southern analysis to confirm differential methylation of these regions. Typical reagents (e.g., as might be found in a typical MCA-based kit) for MCA analysis may include, but are not limited to: PCR primers for arbitrary priming Genomic DNA; PCR buffers and nucleotides, restriction enzymes and appropriate buffers; gene-hybridization oligos or probes; control hybridization oligos or probes.

HeavyMethyl. The HeavyMethyl techniques is a means for selectively amplifying methylated as opposed to non-methylated DNA (or vice versa). Blocker oligonucleotides specific to either methylated or unmethylated versions of a bisulfite treated target sequence are hybridised to the treated nucleic acids. The sample is then enzymatically amplified, wherein the hybridisation of the blocker oligonucleotides hinders amplification of the nucleic acid strand to which it is bound. Typical reagents (e.g., as might be found in a typical HeavyMethyl-based kit) for HeavyMethyl analysis may include, but are not limited to: methylated or unmethylated blocker oligonucleotides for specific gene (or methylation-altered DNA sequence or CpG island), optimized PCR buffers and deoxynucleotides, and specific probes and primers.

GENOMIC SEQUENCES ACCORDING TO SEQ ID NO: 1 to SEQ ID NO: 4 , AND TREATED VARIANTS THEREOF ACCORDING TO SEQ ID NO: 5 to SEQ ID NO: 20, WERE DETERMINED TO HAVE UTILITY FOR DETECTING OR DISTINGUISHING BETWEEN OR AMONG PROSTATE CELL PROLIFERATIVE DISORDERS. .

The present invention is based upon the analysis of methylation levels within one or more genes taken from the group consisting GSTP1, HISTONE H4, PROSTAGLANDIN E2 RECEPTOR and ORPHAN NUCLEAR RECEPTOR NR5A2 and their regulatory regions and sequences thereof according to Table 5.

Particular embodiments of the present invention provide a novel application of the analysis of methylation levels and/or patterns within said genes and/or sequences that enables a precise detection, characterisation and/or treatment of prostate cell proliferative disorders. Early detection of prostate cell proliferative disorders is directly linked with disease prognosis, and the disclosed method thereby enables the physician and patient to make better and more informed treatment decisions. The methods disclosed according to the invention enable the detection and characterisation of prostate cell proliferative disorders with improved sensitivity and/or specificity of with a sensitivity of greater than 30% and a specificity of greater than 65%.

FURTHER IMPROVEMENTS

The present invention provides novel uses for genomic sequences selected from the group consisting of SEQ ID NO: 1 to SEQ ID NO: 4 . Additional embodiments provide modified variants of SEQ ID NO: 1 to SEQ ID NO: 4 , as well as oligonucleotides and/or PNA-oligomers for analysis of cytosine methylation patterns within SEQ ID NO: 1 to SEQ ID NO: 4 .

An objective of the invention comprises analysis of the methylation state of one or more CpG dinucleotides within at least one of the genomic sequences selected from the group consisting of SEQ ID NO: 1 to SEQ ID NO: 4 and sequences complementary thereto.

The disclosed invention provides treated nucleic acids, derived from genomic SEQ ID NO: 1 to SEQ ID NO 4 , wherein the treatment is suitable to convert at least one unmethylated cytosine base of the genomic DNA sequence to uracil or another base that is detectably dissimilar to cytosine in terms of hybridization. The genomic sequences in question may comprise one, or more, consecutive or random methylated CpG positions. Said treatment preferably comprises use of a reagent selected from the group consisting of bisulfite, hydrogen sulfite, disulfite, and combinations thereof .In a preferred embodiment of the invention, the objective comprises analysis of a modified nucleic acid comprising a sequence of at least 16 contiguous nucleotide bases in length of a sequence selected from the group consisting of SEQ ID NO: 5 to SEQ ID NO: 20, wherein said sequence comprises at least one CpG, TpA or CpA

dinucleotide and sequences complementary thereto. The sequences of SEQ ID NO: 5 to SEQ ID NO: 20 provide modified versions of the nucleic acid according to SEQ ID NO: 1 to SEQ ID NO: 4 , wherein the modification of each genomic sequence results in the synthesis of a nucleic acid having a sequence that is unique and distinct from said genomic sequence as follows. For each sense strand genomic DNA, e.g., SEQ ID NO:1, four converted versions are disclosed. A first version wherein "C".."T," but "CpG" remains "CpG" (i.e., corresponds to case where, for the genomic sequence, all "C" residues of CpG dinucleotide sequences are methylated and are thus not converted); a second version discloses the complement of the disclosed genomic DNA sequence (i.e. antisense strand), wherein "C".."T," but "CpG" remains "CpG" (i.e., corresponds to case where, for all "C" residues of CpG dinucleotide sequences are methylated and are thus not converted). The 'upmethylated' converted sequences of SEQ ID NO: 1 to SEQ ID NO: 4 correspond to SEQ ID NO: 5 to SEQ ID NO: 12. A third chemically converted version of each genomic sequences is provided, wherein "C".."T" for all "C" residues, including those of "CpG" dinucleotide sequences (i.e., corresponds to case where, for the genomic sequences, all "C" residues of CpG dinucleotide sequences are unmethylated); a final chemically converted version of each sequence, discloses the complement of the disclosed genomic DNA sequence (i.e. antisense strand), wherein "C".."T" for all "C" residues, including those of "CpG" dinucleotide sequences (i.e., corresponds to case where, for the complement (antisense strand) of each genomic sequence, all "C" residues of CpG dinucleotide sequences are unmethylated). The 'downmethylated' converted sequences of SEQ ID NO: 1 to SEQ ID NO: 4 correspond to SEQ ID NO: 13 to SEQ ID NO: 20.

120

In an alternative preferred embodiment, such analysis comprises the use of an oligonucleotide or oligomer for detecting the cytosine methylation state within genomic or pretreated (chemically modified) DNA, according to SEQ ID NO: 1 to SEQ ID NO: 20. Said oligonucleotide or oligomer comprising a nucleic acid sequence having a length of at least nine (9) nucleotides which hybridizes, under moderately stringent or stringent conditions (as defined herein above), to a pretreated

nucleic acid sequence according to SEQ ID NO: 5 to SEQ ID NO: 20 and/or sequences complementary thereto, or to a genomic sequence according to SEQ ID NO: 1 to SEQ ID NO: 4 and/or sequences complementary thereto.

Thus, the present invention includes nucleic acid molecules (e.g., oligonucleotides and peptide nucleic acid (PNA) molecules (PNA-oligomers)) that hybridize under moderately stringent and/or stringent hybridization conditions to all or a portion of the sequences SEQ ID NO: 1 to SEQ ID NO: 20, or to the complements thereof. The hybridizing portion of the hybridizing nucleic acids is typically at least 9, 15, 20, 25, 30 or 35 nucleotides in length. However, longer molecules have inventive utility, and are thus within the scope of the present invention.

Preferably, the hybridizing portion of the inventive hybridizing nucleic acids is at least 95%, or at least 98%, or 100% identical to the sequence, or to a portion thereof of SEQ ID NO: 1 to SEQ ID NO: 20, or to the complements thereof.

Hybridizing nucleic acids of the type described herein can be used, for example, as a primer (e.g., a PCR primer), or a diagnostic and/or prognostic probe or primer. Preferably, hybridization of the oligonucleotide probe to a nucleic acid sample is performed under stringent conditions and the probe is 100% identical to the target sequence. Nucleic acid duplex or hybrid stability is expressed as the melting temperature or T_m , which is the temperature at which a probe dissociates from a target DNA. This melting temperature is used to define the required stringency conditions.

For target sequences that are related and substantially identical to the corresponding sequence of SEQ ID NO: 1 to SEQ ID NO: 4 (such as allelic variants and SNPs), rather than identical, it is useful to first establish the lowest temperature at which only homologous hybridization occurs with a particular concentration of salt (e.g., SSC or SSPE). Then, assuming that 1% mismatching results in a 1°C decrease in the T_m , the temperature of the final wash in the hybridization reaction is reduced accordingly (for example, if sequences having > 95% identity with the probe are sought, the final wash

temperature is decreased by 5°C). In practice, the change in T_m can be between 0.5°C and 1.5°C per 1% mismatch.

Examples of inventive oligonucleotides of length X (in nucleotides), as indicated by polynucleotide positions with reference to, e.g., SEQ ID NO:1, include those corresponding to sets (sense and antisense sets) of consecutively overlapping oligonucleotides of length X , where the oligonucleotides within each consecutively overlapping set (corresponding to a given X value) are defined as the finite set of Z oligonucleotides from nucleotide positions:

n to $(n + (X-1))$;

where $n=1, 2, 3, \dots (Y-(X-1))$;

where Y equals the length (nucleotides or base pairs) of SEQ ID NO:1 (VALUE TO REFLECT LENGTH OF SEQ ID NO:1);

where X equals the common length (in nucleotides) of each oligonucleotide in the set (e.g., $X=20$ for a set of consecutively overlapping 20-mers); and

where the number (Z) of consecutively overlapping oligomers of length X for a given SEQ ID NO of length Y is equal to $Y-(X-1)$. For example $Z = \text{VALUE TO REFLECT LENGTH OF SEQ ID NO:1} - 19 = \text{VALUE TO REFLECT LENGTH OF SEQ ID NO:1}$ for either sense or antisense sets of SEQ ID NO:1, where $X=20$.

Preferably, the set is limited to those oligomers that comprise at least one CpG, TpG or CpA dinucleotide.

Examples of inventive 20-mer oligonucleotides include the following set of VALUE TO REFLECT LENGTH OF SEQ ID NO:1 oligomers (and the antisense set complementary thereto), indicated by polynucleotide positions with reference to SEQ ID NO:1 1-20, 2-21, 3-22, 4-23, 5-24,

Preferably, the set is limited to those oligomers that comprise at least one CpG, TpG or CpA dinucleotide.

The present invention encompasses, for each of SEQ ID NO: 1 to SEQ ID NO: 20 (sense and antisense), multiple consecutively overlapping sets of oligonucleotides or modified oligonucleotides of length X , where, e.g., $X = 9, 10, 17, 20, 22, 23, 25, 27, 30$ or 35 nucleotides.

The oligonucleotides or oligomers according to the present invention constitute effective tools useful to ascertain

genetic and epigenetic parameters of the genomic sequence corresponding to SEQ ID NO: 1 to SEQ ID NO: 4 . Preferred sets of such oligonucleotides or modified oligonucleotides of length X are those consecutively overlapping sets of oligomers corresponding to SEQ ID NO: 1 to SEQ ID NO: 20 (and to the complements thereof). Preferably, said oligomers comprise at least one CpG, TpG or CpA dinucleotide.

Particularly preferred oligonucleotides or oligomers according to the present invention are those in which the cytosine of the CpG dinucleotide (or of the corresponding converted TpG or CpA dinucleotide) sequences is within the middle third of the oligonucleotide; that is, where the oligonucleotide is, for example, 13 bases in length, the CpG, TpG or CpA dinucleotide is positioned within the fifth to ninth nucleotide from the 5'-end.

The oligonucleotides of the invention can also be modified by chemically linking the oligonucleotide to one or more moieties or conjugates to enhance the activity, stability or detection of the oligonucleotide. Such moieties or conjugates include chromophores, fluorophors, lipids such as cholesterol, cholic acid, thioether, aliphatic chains, phospholipids, polyamines, polyethylene glycol (PEG), palmityl moieties, and others as disclosed in, for example, United States Patent Numbers 5,514,758, 5,565,552, 5,567,810, 5,574,142, 5,585,481, 5,587,371, 5,597,696 and 5,958,773. The probes may also exist in the form of a PNA (peptide nucleic acid) which has particularly preferred pairing properties. Thus, the oligonucleotide may include other appended groups such as peptides, and may include hybridization-triggered cleavage agents (Krol et al., *BioTechniques* 6:958-976, 1988) or intercalating agents (Zon, *Pharm. Res.* 5:539-549, 1988). To this end, the oligonucleotide may be conjugated to another molecule, e.g., a chromophore, fluorophor, peptide, hybridization-triggered cross-linking agent, transport agent, hybridization-triggered cleavage agent, etc.

The oligonucleotide may also comprise at least one art-recognized modified sugar and/or base moiety, or may comprise a modified backbone or non-natural internucleoside linkage.

The oligonucleotides or oligomers according to particular embodiments of the present invention are typically used in

'sets,' which contain at least one oligomer for analysis of each of the CpG dinucleotides of genomic sequence SEQ ID NO: 1 to SEQ ID NO: 4 and sequences complementary thereto, or to the corresponding CpG, TpG or CpA dinucleotide within a sequence of the pretreated nucleic acids according to SEQ ID NO: 5 to SEQ ID NO: 20 and sequences complementary thereto. However, it is anticipated that for economic or other factors it may be preferable to analyze a limited selection of the CpG dinucleotides within said sequences, and the content of the set of oligonucleotides is altered accordingly.

Therefore, in particular embodiments, the present invention provides a set of at least two (2) (oligonucleotides and/or PNA-oligomers) useful for detecting the cytosine methylation state in pretreated genomic DNA (SEQ ID NO: 5 to SEQ ID NO: 20), or in genomic DNA (SEQ ID NO: 1 to SEQ ID NO: 4 and sequences complementary thereto). These probes enable diagnosis, classification and/or therapy of genetic and epigenetic parameters of prostate cell proliferative disorders. The set of oligomers may also be used for detecting single nucleotide polymorphisms (SNPs) in pretreated genomic DNA (SEQ ID NO: 5 to SEQ ID NO: 20), or in genomic DNA (SEQ ID NO: 1 to SEQ ID NO: 4 and sequences complementary thereto).

In preferred embodiments, at least one, and more preferably all members of a set of oligonucleotides is bound to a solid phase.

In further embodiments, the present invention provides a set of at least two (2) oligonucleotides that are used as 'primer' oligonucleotides for amplifying DNA sequences of one of SEQ ID NO: 1 to SEQ ID NO: 20 and sequences complementary thereto, or segments thereof.

It is anticipated that the oligonucleotides may constitute all or part of an "array" or "DNA chip" (i.e., an arrangement of different oligonucleotides and/or PNA-oligomers bound to a solid phase). Such an array of different oligonucleotide- and/or PNA-oligomer sequences can be characterized, for example, in that it is arranged on the solid phase in the form of a rectangular or hexagonal lattice. The solid-phase surface may be composed of silicon, glass, polystyrene, aluminum, steel, iron, copper, nickel, silver, or gold. Nitrocellulose as well as plastics such as nylon, which can exist in the form

of pellets or also as resin matrices, may also be used. An overview of the Prior Art in oligomer array manufacturing can be gathered from a special edition of Nature Genetics (*Nature Genetics Supplement*, Volume 21, January 1999, and from the literature cited therein). Fluorescently labeled probes are often used for the scanning of immobilized DNA arrays. The simple attachment of Cy3 and Cy5 dyes to the 5'-OH of the specific probe are particularly suitable for fluorescence labels. The detection of the fluorescence of the hybridized probes may be carried out, for example, via a confocal microscope. Cy3 and Cy5 dyes, besides many others, are commercially available.

It is particularly preferred that the oligomers according to the invention are utilised for at least one of: detection of; detection and differentiation between or among subclasses of; diagnosis of; prognosis of; treatment of; monitoring of; and treatment and monitoring of prostate cell proliferative disorders. This is enabled by use of said sets for the detection or detection and differentiation of prostate cell proliferative disorders.

The present invention further provides a method for ascertaining genetic and/or epigenetic parameters of the genes GSTP1, HISTONE H4, PROSTAGLANDIN E2 RECEPTOR and ORPHAN NUCLEAR RECEPTOR NR5A2 and their regulatory regions including genomic sequences according to SEQ ID NO: 1 to SEQ ID NO: 4 within a subject by analyzing cytosine methylation and single nucleotide polymorphisms. Said method comprising contacting a nucleic acid comprising one or more of the genes GSTP1, HISTONE H4, PROSTAGLANDIN E2 RECEPTOR and ORPHAN NUCLEAR RECEPTOR NR5A2 and their regulatory regions including genomic sequences according to SEQ ID NO: 1 to SEQ ID NO: 4 in a biological sample obtained from said subject with at least one reagent or a series of reagents, wherein said reagent or series of reagents, distinguishes between methylated and non-methylated CpG dinucleotides within the target nucleic acid. Preferably, said method comprises the following steps: In the first step, a sample of the tissue to be analysed is obtained. The source may be any suitable source, such as cell lines, histological slides, biopsies, tissue embedded in paraffin,

bodily fluids, ejaculate, urine, blood and all possible combinations thereof. The DNA is then isolated from the sample. Extraction may be by means that are standard to one skilled in the art, including the use of commercially available kits, detergent lysates, sonification and vortexing with glass beads. Once the nucleic acids have been extracted, the genomic double stranded DNA is used in the analysis.

In the second step of the method, the genomic DNA sample is treated in such a manner that cytosine bases which are unmethylated at the 5'-position are converted to uracil, thymine, or another base which is dissimilar to cytosine in terms of hybridization behavior. This will be understood as 'pretreatment' herein.

The above described treatment of genomic DNA is preferably carried out with bisulfite (hydrogen sulfite, disulfite) and subsequent alkaline hydrolysis which results in a conversion of non-methylated cytosine nucleobases to uracil or to another base which is dissimilar to cytosine in terms of base pairing behavior.

In the third step of the method, fragments of the pretreated DNA are amplified, using sets of primer oligonucleotides according to the present invention, and an amplification enzyme. The amplification of several DNA segments can be carried out simultaneously in one and the same reaction vessel. Typically, the amplification is carried out using a polymerase chain reaction (PCR). The set of primer oligonucleotides includes at least two oligonucleotides whose sequences are each reverse complementary, identical, or hybridize under stringent or highly stringent conditions to an at least 16-base-pair long segment of the base sequences of one or more of SEQ ID NO: 5 to SEQ ID NO: 20 and sequences complementary thereto.

In an alternate embodiment of the method, the methylation status of preselected CpG positions within the nucleic acid sequences comprising one or more of SEQ ID NO: 1 to SEQ ID NO: 4 may be detected by use of methylation-specific primer oligonucleotides. This technique (MSP) has been described in United States Patent No. 6,265,171 to Herman. The use of methylation status specific primers for the amplification of bisulfite treated DNA allows the differentiation between

methyated and unmethyated nucleic acids. MSP primers pairs contain at least one primer which hybridizes to a bisulfite treated CpG dinucleotide. Therefore, the sequence of said primers comprises at least one CpG dinucleotide. MSP primers specific for non-methyated DNA contain a "T" at the 3' position of the C position in the CpG. Preferably, therefore, the base sequence of said primers is required to comprise a sequence having a length of at least 9 nucleotides which hybridizes to a pretreated nucleic acid sequence according to one of SEQ ID NO: 5 to SEQ ID NO: 20 and sequences complementary thereto, wherein the base sequence of said oligomers comprises at least one CpG dinucleotide.

A further preferred embodiment of the method comprises the use of *blocker* oligonucleotides. The use of such blocker oligonucleotides has been described by Yu et al., *BioTechniques* 23:714-720, 1997. Blocking probe oligonucleotides are hybridized to the bisulfite treated nucleic acid concurrently with the PCR primers. PCR amplification of the nucleic acid is terminated at the 5' position of the blocking probe, such that amplification of a nucleic acid is suppressed where the complementary sequence to the blocking probe is present. The probes may be designed to hybridize to the bisulfite treated nucleic acid in a methylation status specific manner. For example, for detection of methyated nucleic acids within a population of unmethyated nucleic acids, suppression of the amplification of nucleic acids which are unmethyated at the position in question would be carried out by the use of blocking probes comprising a 'CpA' or 'TpA' at the position in question, as opposed to a 'CpG' if the suppression of amplification of methyated nucleic acids is desired.

For PCR methods using blocker oligonucleotides, efficient disruption of polymerase-mediated amplification requires that blocker oligonucleotides not be elongated by the polymerase. Preferably, this is achieved through the use of blockers that are 3'-deoxyoligonucleotides, or oligonucleotides derivitized at the 3' position with other than a "free" hydroxyl group. For example, 3'-O-acetyl oligonucleotides are representative of a preferred class of blocker molecule.

Additionally, polymerase-mediated decomposition of the blocker oligonucleotides should be precluded. Preferably, such

preclusion comprises either use of a polymerase lacking 5'-3' exonuclease activity, or use of modified blocker oligonucleotides having, for example, thioate bridges at the 5'-termini thereof that render the blocker molecule nuclease-resistant. Particular applications may not require such 5' modifications of the blocker. For example, if the blocker- and primer-binding sites overlap, thereby precluding binding of the primer (e.g., with excess blocker), degradation of the blocker oligonucleotide will be substantially precluded. This is because the polymerase will not extend the primer toward, and through (in the 5'-3' direction) the blocker—a process that normally results in degradation of the hybridized blocker oligonucleotide.

A particularly preferred blocker/PCR embodiment, for purposes of the present invention and as implemented herein, comprises the use of peptide nucleic acid (PNA) oligomers as blocking oligonucleotides. Such PNA blocker oligomers are ideally suited, because they are neither decomposed nor extended by the polymerase.

Preferably, therefore, the base sequence of said *blocking oligonucleotides* is required to comprise a sequence having a length of at least 9 nucleotides which hybridizes to a pretreated nucleic acid sequence according to one of SEQ ID NO: 5 to SEQ ID NO: 20 and sequences complementary thereto, wherein the base sequence of said oligonucleotides comprises at least one CpG, TpG or CpA dinucleotide.

The fragments obtained by means of the amplification can carry a directly or indirectly detectable label. Preferred are labels in the form of fluorescence labels, radionuclides, or detachable molecule fragments having a typical mass which can be detected in a mass spectrometer. Where said labels are mass labels, it is preferred that the labeled amplicates have a single positive or negative net charge, allowing for better detectability in the mass spectrometer. The detection may be carried out and visualized by means of, e.g., matrix assisted laser desorption/ionization mass spectrometry (MALDI) or using electron spray mass spectrometry (ESI).

Matrix Assisted Laser Desorption/Ionization Mass Spectrometry (MALDI-TOF) is a very efficient development for

the analysis of biomolecules (Karas & Hillenkamp, *Anal Chem.*, 60:2299-301, 1988). An analyte is embedded in a light-absorbing matrix. The matrix is evaporated by a short laser pulse thus transporting the analyte molecule into the vapour phase in an unfragmented manner. The analyte is ionized by collisions with matrix molecules. An applied voltage accelerates the ions into a field-free flight tube. Due to their different masses, the ions are accelerated at different rates. Smaller ions reach the detector sooner than bigger ones. MALDI-TOF spectrometry is well suited to the analysis of peptides and proteins. The analysis of nucleic acids is somewhat more difficult (Gut & Beck, *Current Innovations and Future Trends*, 1:147-57, 1995). The sensitivity with respect to nucleic acid analysis is approximately 100-times less than for peptides, and decreases disproportionally with increasing fragment size. Moreover, for nucleic acids having a multiply negatively charged backbone, the ionization process via the matrix is considerably less efficient. In MALDI-TOF spectrometry, the selection of the matrix plays an eminently important role. For desorption of peptides, several very efficient matrixes have been found which produce a very fine crystallisation. There are now several responsive matrixes for DNA, however, the difference in sensitivity between peptides and nucleic acids has not been reduced. This difference in sensitivity can be reduced, however, by chemically modifying the DNA in such a manner that it becomes more similar to a peptide. For example, phosphorothioate nucleic acids, in which the usual phosphates of the backbone are substituted with thiophosphates, can be converted into a charge-neutral DNA using simple alkylation chemistry (Gut & Beck, *Nucleic Acids Res.* 23: 1367-73, 1995). The coupling of a charge tag to this modified DNA results in an increase in MALDI-TOF sensitivity to the same level as that found for peptides. A further advantage of charge tagging is the increased stability of the analysis against impurities, which makes the detection of unmodified substrates considerably more difficult.

In the fourth step of the method, the amplificates obtained during the third step of the method are analysed in order to ascertain the methylation status of the CpG dinucleotides prior to the treatment.

In embodiments where the amplificates were obtained by means of MSP amplification, the presence or absence of an amplificate is in itself indicative of the methylation state of the CpG positions covered by the primer, according to the base sequences of said primer.

Amplificates obtained by means of both standard and methylation specific PCR may be further analyzed by means of hybridization-based methods such as, but not limited to, array technology and probe based technologies as well as by means of techniques such as sequencing and template directed extension.

In one embodiment of the method, the amplificates synthesised in *step three* are subsequently hybridized to an array or a set of oligonucleotides and/or PNA probes. In this context, the hybridization takes place in the following manner: the set of probes used during the hybridization is preferably composed of at least 2 oligonucleotides or PNA-oligomers; in the process, the amplificates serve as probes which hybridize to oligonucleotides previously bonded to a solid phase; the non-hybridized fragments are subsequently removed; said oligonucleotides contain at least one base sequence having a length of at least 9 nucleotides which is reverse complementary or identical to a segment of the base sequences specified in the present Sequence Listing; and the segment comprises at least one CpG , TpG or CpA dinucleotide.

In a preferred embodiment, said dinucleotide is present in the central third of the oligomer. For example, wherein the oligomer comprises one CpG dinucleotide, said dinucleotide is preferably the fifth to ninth nucleotide from the 5'-end of a 13-mer. One oligonucleotide exists for the analysis of each CpG dinucleotide within the sequence according to SEQ ID NO: 1 to SEQ ID NO: 4 , and the equivalent positions within SEQ ID NO: 5 to SEQ ID NO: 20. Said oligonucleotides may also be present in the form of peptide nucleic acids. The non-hybridized amplificates are then removed. The hybridized amplificates are then detected. In this context, it is preferred that labels attached to the amplificates are identifiable at each position of the solid phase at which an oligonucleotide sequence is located.

In yet a further embodiment of the method, the genomic methylation status of the CpG positions may be ascertained by

means of oligonucleotide probes that are hybridised to the bisulfite treated DNA concurrently with the PCR amplification primers (wherein said primers may either be methylation specific or standard).

A particularly preferred embodiment of this method is the use of fluorescence-based Real Time Quantitative PCR (Heid et al., *Genome Res.* 6:986-994, 1996; also see United States Patent No. 6,331,393) employing a dual-labeled fluorescent oligonucleotide probe (TaqMan™ PCR, using an ABI Prism 7700 Sequence Detection System, Perkin Elmer Applied Biosystems, Foster City, California). The TaqMan™ PCR reaction employs the use of a nonextendible interrogating oligonucleotide, called a TaqMan™ probe, which, in preferred imbodiments, is designed to hybridize to a GpC-rich sequence located between the forward and reverse amplification primers. The TaqMan™ probe further comprises a fluorescent "reporter moiety" and a "quencher moiety" covalently bound to linker moieties (e.g., phosphoramidites) attached to the nucleotides of the TaqMan™ oligonucleotide. For analysis of methylation within nucleic acids subsequent to bisulfite treatment, it is required that the probe be methylation specific, as described in United States Patent No. 6,331,393, (hereby incorporated by reference in its entirety) also known as the MethyLight™ assay. Variations on the TaqMan™ detection methodology that are also suitable for use with the described invention include the use of dual-probe technology (Lightcycler™) or fluorescent amplification primers (Sunrise™ technology). Both these techniques may be adapted in a manner suitable for use with bisulfite treated DNA, and moreover for methylation analysis within CpG dinucleotides.

A further suitable method for the use of probe oligonucleotides for the assessment of methylation by analysis of bisulfite treated nucleic acids In a further preferred embodiment of the method, the *fifth step* of the method comprises the use of template-directed oligonucleotide extension, such as MS-SNuPE as described by Gonzalgo & Jones, *Nucleic Acids Res.* 25:2529-2531, 1997.

In yet a further embodiment of the method, the *fifth step* of the method comprises sequencing and subsequent sequence analysis of the amplificate generated in the *third step* of the

method (Sanger F., et al., *Proc Natl Acad Sci USA* 74:5463-5467, 1977).

Best mode

In the most preferred embodiment of the method the nucleic acids according to SEQ ID NO: 1 to SEQ ID NO 4 are isolated and treated according to the first three steps of the method outlined above, namely:

- a. obtaining, from a subject, a biological sample having subject genomic DNA;
- b. extracting or otherwise isolating the genomic DNA;
- c. treating the genomic DNA of b), or a fragment thereof, with one or more reagents to convert cytosine bases that are unmethylated in the 5-position thereof to uracil or to another base that is detectably dissimilar to cytosine in terms of hybridization properties;

and wherein the subsequent amplification of d) is carried out in a methylation specific manner, namely by use of methylation specific primers or *blocking oligonucleotides*, and further wherein the detection of the amplicates is carried out by means of a real-time detection probes, as described above.

Wherein the subsequent amplification of d) is carried out by means of methylation specific primers, as described above, said methylation specific primers comprise a sequence having a length of at least 9 nucleotides which hybridizes to a pretreated nucleic acid sequence according to one of SEQ ID NO: 5 to SEQ ID NO: 20 and sequences complementary thereto, wherein the base sequence of said oligomers comprises at least one CpG dinucleotide. Step e) of the method, namely the detection of the specific amplicates indicative of the methylation status of one or more CpG positions according to SEQ ID NO: 1 to SEQ ID NO 4 is carried out by means of real-time detection methods as described above.

In an alternative most preferred embodiment of the method the subsequent amplification of d) is carried out in the presence of *blocking oligonucleotides*, as described above. Said *blocking oligonucleotides* comprising a sequence having a length of at least 9 nucleotides which hybridizes to a pretreated nucleic acid sequence according to one of SEQ ID NO: 5 to SEQ ID NO: 20

and sequences complementary thereto, wherein the base sequence of said oligomers comprises at least one CpG, TpG or CpA dinucleotide. Step e) of the method, namely the detection of the specific amplicates indicative of the methylation status of one or more CpG positions according to SEQ ID NO: 1 to SEQ ID NO 4 is carried out by means of real-time detection methods as described above.

14 14 14

Diagnostic and/or Prognostic Assays for prostate cell proliferative disorders

The present invention enables diagnosis of events which are disadvantageous to patients or individuals in which important genetic and/or epigenetic parameters within one or more of the genes GSTP1, HISTONE H4, PROSTAGLANDIN E2 RECEPTOR and ORPHAN NUCLEAR RECEPTOR NR5A2 and their regulatory regions including genomic sequences according to SEQ ID NO: 1 to SEQ ID NO: 4 may be used as markers. Said parameters obtained by means of the present invention may be compared to another set of genetic and/or epigenetic parameters, the differences serving as the basis for a diagnosis and/or prognosis of events which are disadvantageous to patients or individuals.

Specifically, the present invention provides for diagnostic and/or prognostic cancer assays based on measurement of differential methylation of one or more CpG dinucleotide sequences of the genes GSTP1, HISTONE H4, PROSTAGLANDIN E2 RECEPTOR and ORPHAN NUCLEAR RECEPTOR NR5A2 and their regulatory regions including genomic sequences according to SEQ ID NO: 1 to SEQ ID NO: 4, or of subregions thereof that comprise such a CpG dinucleotide sequence. Typically, such assays involve obtaining a tissue sample from a test tissue, performing an assay to measure the methylation status of at least one of one or more CpG dinucleotide sequences of the genes GSTP1, HISTONE H4, PROSTAGLANDIN E2 RECEPTOR and ORPHAN NUCLEAR RECEPTOR NR5A2 and their regulatory regions including genomic sequences according to SEQ ID NO: 1 to SEQ ID NO: 4 derived from the tissue sample, relative to a control sample, or a known standard and making a diagnosis or prognosis based thereon.

In particular preferred embodiments, inventive oligomers are used to assess the CpG dinucleotide methylation status, such as those based on SEQ ID NO: 1 to SEQ ID NO: 20, or arrays thereof, as well as in kits based thereon and useful for the diagnosis and/or prognosis of prostate cell proliferative disorders.

Kits

Moreover, an additional aspect of the present invention is a kit comprising, for example: a bisulfite-containing reagent; a set of primer oligonucleotides containing at least two oligonucleotides whose sequences in each case correspond, are complementary, or hybridize under stringent or highly stringent conditions to a 16-base long segment of one or more of the genes GSTP1, HISTONE H4, PROSTAGLANDIN E2 RECEPTOR and ORPHAN NUCLEAR RECEPTOR NR5A2 and their regulatory regions including genomic and/or treated sequences according to SEQ ID NO: 1 to SEQ ID NO: 20; oligonucleotides and/or PNA-oligomers; as well as instructions for carrying out and evaluating the described method. In a further preferred embodiment, said kit may further comprise standard reagents for performing a CpG position-specific methylation analysis, wherein said analysis comprises one or more of the following techniques: MS-SNuPE, MSP, MethyLight™, HeavyMethyl™, COBRA, and nucleic acid sequencing. However, a kit along the lines of the present invention can also contain only part of the aforementioned components.

While the present invention has been described with specificity in accordance with certain of its preferred embodiments, the following example serves only to illustrate the invention and is not intended to limit the invention within the principles and scope of the broadest interpretations and equivalent configurations thereof.

EXAMPLES

The objective of the following study was to analyze the methylation status of prostate cancer markers in different body fluid samples in order to identify the preferred choice of body

fluid (urine or serum) for testing and the preferred marker, markers or combinations of markers. The study was run on matched serum and urine sediment samples from 80 patients with an average age of 65 and representative of a number of racial types (caucasian, african american etc..). In each case, genomic DNA was analyzed using the HeavyMethyl or MSP technique after bisulfite conversion.

Urine Sediment was prepared for analysis and bisulphite treated according to the following:

- 200 ul sediment samples were purified using the Magnapure DNA Isolation Kit 1 with a 100 ul elution volume.
- 5 ul HD6 PCR was carried out on the Magnapure Eluate , in order to determine DNA concentration
- 100 ul of the DNA solution was treated using a proprietary bisulfite treatment technique
- 10 ul C3 bisulfite specific quantitative PCR
- 5 ul Merck sulfite test

Serum was prepared for analysis and bisulphite treated according to the following:

- 1 mL serum samples were purified using the Magnapure DNA Large Volume Total nucleic acid with a 100 ul elution volume.
- 5 ul HD6 PCR on Magnapure Eluate - To determine DNA concentration
- 100 ul of the DNA solution was treated using a proprietary bisulfite treatment technique
- 10 ul C3 bisulfite specific quantitative PCR
- 5 ul Merck sulfite test

Single PCR runs were performed on 10 ul of bisulfite treated DNA per sample for each of the markers as described below.

Heavy Methyl Assay of the GSTP1 gene

In the following analysis the methylation status of the gene GSTP1 was analysed by means of methylation specific amplification using the primers according to Table 1 (below).

The sequence of interest is amplified by means of methylation specific primers and a blocker oligonucleotide in order to minimise the unspecific amplification of non methylated DNA. The amplificate is then detected by means of methylation specific Lightcycler probes.

Table 1: Oligonucleotides for MSP - Lightcycler analysis of GSTP1.

SEQ ID NO:	Sequence	Type
21	gggattatTTTTtataagggtt	primer
22	ctctaaaccccatcccc	primer
23	cccatccccaaaaacacaaaccac	blocker
24	CGtCGtCGtAGTtTTCGtt-fluo	probe
25	red640-tAGTGAGTACGCGCGGtt-pho	probe

Reaction conditions:

PCR program

denat at 95°C

95°C 10min

50 cycles: ramp

denat at 95°C 10 sec (1°C/s)

annealing 56°C 30 sec (1°C/s) detection

extension 72°C 10 sec (1°C/s)

MSP analysis of the gene HISTONE H4.

In the following analysis the methylation status of the gene HISTONE H4 was analysed by means of methylation specific

amplification using the primers according to Table 2 (below).

The sequence of interest is amplified by means of methylation specific primers, the amplificate is then detected by means of methylation specific Taqman probes.

Table 2: Oligonucleotides for MSP - Taqman analysis of HISTONE H4.

SEQ ID NO :	Sequence	Type
26	accgaaaatacgcttcacg	primer
27	gcgttatcgtaaagtattgcgc	primer
28	/56-FAM/cgcgacgaacaaaacgccg/3BHQ_1/	probe

Reaction Conditions:

PCR program

denat at 95°C

95°C 10min

50 cycles: ramp

denat at 95°C 10 sec (20°C/s)

annealing 60°C 45 sec (20°C/s) detection

MSP analysis of the gene PROSTAGLANDIN E2 RECEPTOR.

In the following analysis the methylation status of the gene PROSTAGLANDIN E2 RECEPTOR was analysed by means of methylation specific amplification using the primers according to Table 3 (below).

The sequence of interest is amplified by means of methylation specific primers, the amplificate is then detected by means of methylation specific Taqman probes.

Table 3: Oligonucleotides for MSP - Taqman analysis of PROSTAGLANDIN E2 RECEPTOR

SEQ ID NO :	Sequence	Type
29	cgcgctactccgcataca	primer
30	gaggtaatcgaggcggtcg	primer
31	/56-FAM/cgccaattcatagccgcacc/3BHQ_1/	probe

PCR programdenat at 95°C

95°C 10min

50 cycles: ramp

denat at 95°C 10 sec (20°C/s)

annealing 60°C 45 sec (20°C/s) detection

MSP analysis of the gene ORPHAN NUCLEAR RECEPTOR NR5A2.

In the following analysis the methylation status of the gene ORPHAN NUCLEAR RECEPTOR NR5A2 was analysed by means of methylation specific amplification using the primers according to Table 3 (below).

The sequence of interest is amplified by means of methylation specific primers, the amplificate is then detected by means of methylation specific Taqman probes.

Table 4: Oligonucleotides for MSP - Taqman analysis of ORPHAN NUCLEAR RECEPTOR NR5A2.

SEQ ID NO :	Sequence	Type
32	ttgtggttcgggaagagac	primer
33	tcccgaaactcttcgatcg	primer
34	aactacgcgcaaaccgcga	probe

PCR program

denat at 95°C

95°C 10min

50 cycles: ramp

denat at 95°C 10 sec (20°C/s)

annealing 60°C 45 sec (20°C/s) detection

Marker Analysis

Results were analyzed qualitatively by scoring amplification as +/- and quantitatively by determining the percentage of methylated DNA as a fraction of total DNA calculated using the C3 bisulfite specific PCR. To measure total methylated DNA, a 100% methylated standard (chemicon SSS1 treated DNA) standard curve was included in each assay.

Results

For each marker a Receiver Operating Characteristic curve (ROC curve) of the assay was determined. A ROC is a plot of the true positive rate against the false positive rate for the different possible cutpoints of a diagnostic test. It shows the tradeoff between sensitivity and specificity depending on the selected cutpoint (any increase in sensitivity will be accompanied by a decrease in specificity). The area under an ROC curve (AUC) is a measure for the accuracy of a diagnostic test (the larger the area the better, optimum is 1, a random test would have a ROC curve lying on the diagonal with an area of 0.5; for reference: J.P. Egan. Signal Detection Theory and ROC Analysis, Academic Press, New York, 1975).

AUC results:

Serum:

Marker: HevyMethyl GSTP1

AUC: 0.51

Marker: MSP HISTONE H4

AUC: 0.59

Marker: MSP PROSTAGLANDIN E2 RECEPTOR

AUC: 0.52

Marker: MSP ORPHAN NUCLEAR RECEPTOR NR5A2

AUC: 0.50

Urine:

Marker: HeavyMethyl GSTP1

AUC: 0.58

Marker: MSP HISTONE H4

AUC: 0.5

Marker: MSP PROSTAGLANDIN E2 RECEPTOR

AUC: 0.49

Marker: MSP ORPHAN NUCLEAR RECEPTOR NR5A2

AUC: 0.56

In order to provide an accurate detection of prostate cancer it is preferred that a combined analysis of multiple markers is carried out (i.e. a gene panel). For analysis of urine based samples the most preferred combination of markers is GSTP1 , PROSTAGLANDIN E2 RECEPTOR & ORPHAN NUCLEAR RECEPTOR NR5A2 with a sensitivity of 0.37 and a specificity of 0.72.

For analysis of serum based samples the most preferred combination of markers is GSTP1 , HISTONE H4 & ORPHAN NUCLEAR RECEPTOR NR5A2 with a sensitivity of 0.35 and a specificity of 0.75.

Table 5: Genes & Sequences according to the invention

Gene Name	Genbank Ref. Seq.	Genomic sequence SEQ ID NO:	Treated sequences (methylated) SEQ ID NO:	Treated sequences (unmethylated) SEQ ID NO:
GSTP1	NM_000852	1	5 & 6	13 & 14
HISTONE H4	NM_003495	2	7 & 8	15 & 16
PROSTAGLANDIN E2 RECEPTOR	NM_000958	3	9 & 10	17 & 18
ORPHAN NUCLEAR RECEPTOR NR5A2	NM_003822	4	11 & 12	19 & 20

10 -02- 2004

We claim:

1. A method for detecting, or for detecting and distinguishing between or among prostate cell proliferative disorders in a subject with a sensitivity of greater than 30% and a specificity of greater than 65%, said method comprising analysing the methylation pattern of a target nucleic acid comprising one or a combination of sequences taken from the group consisting of SEQ ID Nos: 1-4 by contacting at least one of said target nucleic acids in a biological sample obtained from said subject with at least one reagent, or series of reagents that distinguishes between methylated and non-methylated CpG dinucleotides.
2. The method of claim 1, wherein prostate carcinoma is distinguished from at least one condition selected from the group consisting of prostate adenoma, normal prostate tissue, non-prostate tissues and non-prostate cell proliferative disorders.
3. A method according to claim 1, comprising:
 - obtaining, from a subject, a biological sample having subject genomic DNA;
 - contacting the genomic DNA, or a fragment thereof, with one reagent or a plurality of reagents for distinguishing between methylated and non methylated CpG dinucleotide sequences within at least one target sequence of the genomic DNA, or fragment thereof, wherein the target sequence comprises, or hybridizes under stringent conditions to, at least 16 contiguous nucleotides of a sequence taken from the group consisting of SEQ ID NO: 1 to SEQ ID NO 4, said contiguous nucleotides comprising at least one CpG dinucleotide sequence; and
 - determining, based at least in part on said distinguishing, the methylation state of at least one target CpG dinucleotide sequence, or an average, or a value reflecting an average methylation state of a plurality of target CpG dinucleotide sequences, whereby detecting, or detecting and distinguishing between or

among prostate cell proliferative disorders with a sensitivity of greater than 30% and a specificity of greater than 65% is, at least in part, afforded.

4. The method of claim 3, wherein distinguishing between methylated and non methylated CpG dinucleotide sequences within the target sequence comprises converting unmethylated cytosine bases within the target sequence to uracil or to another base that is detectably dissimilar to cytosine in terms of hybridization properties.
5. The method of claim 3, wherein distinguishing between methylated and non methylated CpG dinucleotide sequences within the target sequence(s) comprises methylation state-dependent conversion or non-conversion of at least one CpG dinucleotide sequence to the corresponding converted or non-converted dinucleotide sequence within a sequence selected from the group consisting of SEQ ID NO: 5 to SEQ ID NO 20, and contiguous regions thereof corresponding to the target sequence.
6. The method of claim 3, wherein the biological sample is selected from the group consisting of cell lines, histological slides, biopsies, paraffin-embedded tissue, bodily fluids, ejaculate, urine, blood, and combinations thereof.
7. The method of claim 3, wherein distinguishing between methylated and non methylated CpG dinucleotide sequences within the target sequence comprises use of at least one nucleic acid molecule or peptide nucleic acid (PNA) molecule comprising, in each case a contiguous sequence at least 9 nucleotides in length that is complementary to, or hybridizes under moderately stringent or stringent conditions to a sequence selected from the group consisting of SEQ ID NO: 5 to SEQ ID NO 20, and complements thereof.
8. The method of claim 7, wherein the contiguous sequence comprises at least one CpG, TpG or CpA dinucleotide sequence.

9. The method of claim 7, comprising use of at least two such nucleic acid molecules, or peptide nucleic acid (PNA) molecules.

10. The method of claim 7, comprising use of at least two such nucleic acid molecules, or peptide nucleic acid (PNA) molecules as primer oligonucleotides for the amplification of a sequences selected from the group consisting of SEQ ID NO: 5 to SEQ ID NO 20, sequences complementary thereto, and regions thereof that comprise, or hybridize under stringent conditions to the primers.

11. The method of claim 9, comprising use of at least four such nucleic acid molecules, or peptide nucleic acid (PNA) molecules.

12. A method for detecting, or detecting and distinguishing between or among prostate cell proliferative disorders in a subject, comprising:

- a. obtaining, from a subject, a biological sample having subject genomic DNA;
- b. extracting or otherwise isolating the genomic DNA;
- c. treating the genomic DNA of b), or a fragment thereof, with one or more reagents to convert cytosine bases that are unmethylated in the 5-position thereof to uracil or to another base that is detectably dissimilar to cytosine in terms of hybridization properties;
- d. contacting the treated genomic DNA, or the treated fragment thereof, with an amplification enzyme and at least two primers comprising, in each case a contiguous sequence of at least 9 nucleotides that is complementary to, or hybridizes under moderately stringent or stringent conditions to a sequence selected from the group consisting of SEQ ID NO: 5 to SEQ ID NO 20, and complements thereof, wherein the treated genomic DNA or the fragment thereof is either amplified to produce at least one amplificate, or is not amplified; and

e) determining, based on a presence or absence of, or on a property of said amplificate, the methylation state of at least one CpG dinucleotide of a sequence selected from the group consisting SEQ ID NO: 1 to SEQ ID NO 4 , or an average, or a value reflecting an average methylation state of a plurality of CpG dinucleotides of a sequence selected from the groups consisting of SEQ ID NO: 1 to SEQ ID NO 4 , whereby at least one of detecting, or detecting and distinguishing between prostate cell proliferative disorders with a sensitivity of greater than 30% and a specificity of greater than 65% is at least in part, afforded.

13. The method of claim 12, wherein treating the genomic DNA, or the fragment thereof in c), comprises use of a reagent selected from the group consisting of bisulfite, hydrogen sulfite, disulfite, and combinations thereof.
14. The method of claim 12, wherein contacting or amplifying in d) comprises use of at least one method selected from the group consisting of: use of a heat-resistant DNA polymerase as the amplification enzyme; use of a polymerase lacking 5'-3' exonuclease activity; use of a polymerase chain reaction (PCR); generation of a amplificate nucleic acid molecule carrying a detectable labels; and combinations thereof.
15. The method of claim 14, wherein the detectable amplificate label is selected from the label group consisting of: fluorescent labels; radionuclides or radiolabels; amplificate mass labels detectable in a mass spectrometer; detachable amplificate fragment mass labels detectable in a mass spectrometer; amplificate, and detachable amplificate fragment mass labels having a single-positive or single-negative net charge detectable in a mass spectrometer; and combinations thereof.

16. The method of claim 12, wherein the biological sample obtained from the subject is selected from the group consisting of cell lines, histological slides, biopsies, paraffin-embedded tissue, bodily fluids, ejaculate, urine, blood, and combinations thereof.
17. The method of claim 12, wherein prostate carcinoma is distinguished from at least one condition selected from the group consisting of prostate adenoma, inflammatory prostate tissue, prostate adenomas with grade 2 dysplasia less than 1 cm, prostate adenomas with grade 3 dysplasia equal to or greater than 1 cm in size, normal prostate tissues, non-prostate normal tissue, body fluids, and non-prostate cancer tissue.
18. The method of claim 12, further comprising in step d) the use of at least one nucleic acid molecule or peptide nucleic acid molecule comprising in each case a contiguous sequence at least 9 nucleotides in length that is complementary to, or hybridizes under moderately stringent or stringent conditions to a sequence selected from the group consisting of SEQ ID NO: 5 to SEQ ID NO 20, and complements thereof, wherein said nucleic acid molecule or peptide nucleic acid molecule suppresses amplification of the nucleic acid to which it is hybridized.
19. The method of claim 18, wherein said nucleic acid molecule or peptide nucleic acid molecule is in each case modified at the 5'-end thereof to preclude degradation by an enzyme having 5'-3' exonuclease activity.
20. The method of claim 18, wherein said nucleic acid molecule or peptide nucleic acid molecule is in each case lacking a 3' hydroxyl group.
21. The method of claim 18, wherein the amplification enzyme is a polymerase lacking 5'-3' exonuclease activity.

22. The method of claim 12, wherein determining in e) comprises hybridization of at least one nucleic acid molecule or peptide nucleic acid molecule in each case comprising a contiguous sequence at least 9 nucleotides in length that is complementary to, or hybridizes under moderately stringent or stringent conditions to a sequence selected from the group consisting of SEQ ID NO: 5 to SEQ ID NO: 20, and complements thereof.
23. The method of claim 22, wherein at least one such hybridizing nucleic acid molecule or peptide nucleic acid molecule is bound to a solid phase.
24. The method of claim 22, wherein a plurality of such hybridizing nucleic acid molecules or peptide nucleic acid molecules are bound to a solid phase in the form of a nucleic acid or peptide nucleic acid array selected from the array group consisting of linear, or substantially so, hexagonal or substantially so, rectangular or substantially so, and combinations thereof.
25. The method of claim 22, further comprising extending at least one such hybridized nucleic acid molecule by at least one nucleotide base.
26. The method of claim 12, wherein determining in e), comprises sequencing of the amplificate.
27. The method of claim 12, wherein contacting or amplifying in d), comprises use of methylation-specific primers.
28. The method of claim 12 comprising in d) using primer oligonucleotides comprising one or more CpG; TpG or CpA dinucleotides; and further comprising in e) the use of at least one method selected from the group consisting of: hybridizing in at least one nucleic acid molecule or peptide nucleic acid molecule comprising a contiguous sequence at least 9 nucleotides in length that

is complementary to, or hybridizes under moderately stringent or stringent conditions to a sequence selected from the group consisting of SEQ ID NO: 5 to SEQ ID NO 20, and complements thereof; hybridizing at least one nucleic acid molecule that is bound to a solid phase and comprises a contiguous sequence at least 9 nucleotides in length that is complementary to, or hybridizes under moderately stringent or stringent conditions to a sequence selected from the group consisting of SEQ ID NO: 5 to SEQ ID NO 20, and complements thereof; hybridizing at least one nucleic acid molecule comprising a contiguous sequence at least 9 nucleotides in length that is complementary to, or hybridizes under moderately stringent or stringent conditions to a sequence selected from the group consisting of SEQ ID NO: 5 to SEQ ID NO 20, and complements thereof, and extending at least one such hybridized nucleic acid molecule by at least one nucleotide base; and sequencing-in e) of the amplificate.

29. The method of claim 12 comprising in d) use of at least one nucleic acid molecule or peptide nucleic acid molecule comprising in each case a contiguous sequence at least 9 nucleotides in length that is complementary to, or hybridizes under moderately stringent or stringent conditions to a sequence selected from the group consisting of SEQ ID NO: 5 to SEQ ID NO 20, and complements thereof, wherein said nucleic acid molecule or peptide nucleic acid molecule suppresses amplification of the nucleic acid to which it is hybridized; and further comprising in e) the use of at least one method selected from the group consisting of: hybridizing in at least one nucleic acid molecule or peptide nucleic acid molecule comprising a contiguous sequence at least 9 nucleotides in length that is complementary to, or hybridizes under moderately stringent or stringent conditions to a sequence selected from the group consisting of SEQ ID NO: 5 to SEQ ID NO 20, and complements thereof; hybridizing at least one nucleic acid molecule that is bound to a solid phase and comprises a contiguous sequence at least 9 nucleotides in length that is complementary to, or hybridizes under

moderately stringent or stringent conditions to a sequence selected from the group consisting of SEQ ID NO: 5 to SEQ ID NO 20, and complements thereof; hybridizing at least one nucleic acid molecule comprising a contiguous sequence at least 9 nucleotides in length that is complementary to, or hybridizes under moderately stringent or stringent conditions to a sequence selected from the group consisting of SEQ ID NO: 5 to SEQ ID NO 20, and complements thereof, and extending at least one such hybridized nucleic acid molecule by at least one nucleotide base; and sequencing in e) of the amplificate.

30. The method of claim 12, comprising in d) amplification by primer oligonucleotides comprising one or more CpG; TpG or CpA dinucleotides and further comprising in e) hybridizing at least one detectably labeled nucleic acid molecule comprising a contiguous sequence at least 9 nucleotides in length that is complementary to, or hybridizes under moderately stringent or stringent conditions to a sequence selected from the group consisting of SEQ ID NO: 5 to SEQ ID NO 20.

31. The method of claim 12, comprising in d) the use of at least one nucleic acid molecule or peptide nucleic acid molecule comprising in each case a contiguous sequence at least 9 nucleotides in length that is complementary to, or hybridizes under moderately stringent or stringent conditions to a sequence selected from the group consisting of SEQ ID NO: 5 to SEQ ID NO 20, and complements thereof, wherein said nucleic acid molecule or peptide nucleic acid molecule suppresses amplification of the nucleic acid to which it is hybridized, and further comprising in e) hybridizing at least one detectably labeled nucleic acid molecule comprising a contiguous sequence at least 9 nucleotides in length that is complementary to, or hybridizes under moderately stringent or stringent conditions to a sequence selected from the group consisting of SEQ ID NO: 5 to SEQ ID NO 20.

32. 14 14 14 A treated nucleic acid derived from genomic SEQ ID NO: 1 to SEQ ID NO 4 , wherein the treatment is suitable to convert at least one unmethylated cytosine base of the genomic DNA sequence to uracil or another base that is detectably dissimilar to cytosine in terms of hybridization.
33. A nucleic acid, comprising at least 16 contiguous nucleotides of a treated genomic DNA sequence selected from the group consisting of SEQ ID NO: 5 to SEQ ID NO 20, and sequences complementary thereto, wherein the treatment is suitable to convert at least one unmethylated cytosine base of the genomic DNA sequence to uracil or another base that is detectably dissimilar to cytosine in terms of hybridization.
34. The nucleic acid of claims 32 and 33 wherein the contiguous base sequence comprises at least one CpG, TpG or CpA dinucleotide sequence.
35. The nucleic acid of claims 32 and 33 wherein the treatment comprises use of a reagent selected from the group consisting of bisulfite, hydrogen sulfite, disulfite, and combinations thereof .
36. An oligomer, comprising a sequence of at least 9 contiguous nucleotides that is complementary to, or hybridizes under moderately stringent or stringent conditions to a treated genomic DNA sequence selected from the group consisting of SEQ ID NO: 5 to SEQ ID NO 20.
37. The oligomer of Claim 36, comprising at least one CpG , CpA or TpG dinucleotide sequence.
38. An oligomer of claim 37, having a sequence selected from the group consisting of SEQ ID NO: &[IDOLIGOFIRST] to SEQ ID NO &[IDOLIGOLAST].

39. A set of oligomers, comprising at least two oligonucleotides according, in each case, to any one of Claims 37 or 38.
40. Use of a set of oligomers according, in each case, to any one of Claims 36 through 39, as probes for determining at least one of a cytosine methylation state, or a single nucleotide polymorphism (SNP) of a sequence selected from the group consisting of SEQ ID NO: 1 to 4 and sequences complementary thereto.
- 41.
42. A kit useful for detecting, or for detecting distinguishing between or among prostate cell proliferative disorders of a subject, comprising:
- at least one of a bisulfite reagent, or a methylation-sensitive restriction enzyme; and
 - at least one nucleic acid molecule or peptide nucleic acid molecule comprising, in each case a contiguous sequence at least 9 nucleotides that is complementary to, or hybridizes under moderately stringent or stringent conditions to a sequence selected from the group consisting of SEQ ID 5 to SEQ ID NO 20, and complements thereof
43. The kit of claim 42, further comprising standard reagents for performing a methylation assay selected from the group consisting of MS-SNuPE, MSP, MethyLight, HeavyMethyl, COBRA, nucleic acid sequencing, and combinations thereof.
44. The method of any one of claims 1, 12 or 3 comprising use of the kit according to claim 53.
45. Use of a nucleic acid according to claims 32 through 35, an oligomer according to any one of claims 36 through 38, a set of oligonucleotides according to claim 39 and a kit according to claims 43 and 44 for the detection of,

detection and differentiation between or among subclasses
of prostate cell proliferative disorders.

10-02-2004

Abstract

The invention provides methods, nucleic acids and kits for detecting, or for detecting and distinguishing between or among prostate cell proliferative disorders. The invention discloses genomic sequences the methylation patterns of which have utility for the improved detection of and differentiation between said class of disorders, thereby enabling the improved diagnosis and treatment of patients.

10-02-2004

<110> Epigenomics AG

<120> METHODS AND NUCLEIC ACIDS FOR THE ANALYSIS OF CpG DINUCLEOTIDE METHYLATION STATUS ASSOCIATED WITH THE DEVELOPMENT OF PROSTATE CANCER.

<160> 34

<210> 1

<211> 2501

<212> DNA

<213> Homo Sapiens

<400> 1

ttgtttgtaca	gaatattttca	tcacccaggt	attatgccga	gtacccaata	gttctctttt	60
ctgctcctct	ccttctctcc	atcctgcacc	ctggagtcaa	ccacagtgtc	tggtgtttcc	120
ttgttttgtgt	tataagttct	catcatttag	ctcccactta	caagtgagaa	catccagtat	180
ttggattttct	gttctctgcat	tagtttgcta	aggataatag	cctctagctc	catccatggt	240
cccacaaaag	acatgatcta	gttcttttta	atggctgcat	taaatgaagt	tttaaagata	300
caacataaac	accaacctct	tccccaccac	aaaaatccct	tgctgaattt	gattacactt	360
aaattaacga	gttttgtttc	atgaaagact	ccttggacaa	acttgacagt	tgatggaata	420
ggagaagctg	tctgtcatgt	ctaaaagccaa	caagagatca	atatctagaa	taaatggaga	480
tctgcaaatc	aacagaaagt	aggcagcaaa	gccaaagaaa	atagcctaag	gcacagccac	540
taaaaggaac	gtgatcatgt	cctttgcagg	gacatgggtg	gagctggaag	ccgttagcct	600
cagcaaaactc	acacaggaac	agaaaaccag	cgagaccgca	tggtctcact	tataagtggg	660
agctgaacaa	tgagaacaca	tggtcacatg	gcggcgatca	acacacactg	gtgctgtttg	720
agcgggggtgc	tggggaggga	gagtaccagg	aagaatagct	aagggatact	gggcttaata	780
cctgggtgat	gggatgatct	gtacagcaaa	ccatcatggc	gcacacacct	atgtaacaaa	840
cctgcacatc	ctctacatgt	accccagaac	ttcaaataaa	agttggacgg	ccaggcgtgg	900
tggctcacgc	ctgtaatccc	agcacttttg	gaagccgagg	cgtagcagatc	acctaagggtc	960
aggagttcga	gaccagcccc	gccaacatgg	tgaaaccccc	tctctactaa	aaatacaaaa	1020
atcagccaga	tgtggcacgc	acctataatt	ccacctactc	gggaggctga	agcagaattg	1080
cttgaaccgc	agaggcggag	gttgcagtga	gccgcggaga	tgcgcgcact	gcactccagc	1140
cctgggccaca	gcgtgagact	acgtcataaa	ataaaataaa	ataacacaaa	ataaaataaa	1200
ataaaataaa	ataaaataaa	ataaaataaa	ataaaataaa	ataaaaaaat	aaaataaaaat	1260
aaaataaaaat	aaagcaattt	cctttcctct	aagcggcctc	cacccctctc	ccctgccctg	1320
tgaagcgggt	gtgcaagctc	cgggatcgca	gcggtcttag	ggaatttccc	cccgcgatgt	1380
cccggcgcgc	cagttcgctg	cgcacacttc	gctgcggctc	tcttctgct	gtctgtttac	1440
tccctaggcc	ccgtcgggga	cctgggaaaag	agggaaaagg	ttccccggcc	agctgcgcgg	1500
cgactccggg	gactccaggg	cgccctctg	cgcccgacgc	ccggggtgca	gcggccgcgc	1560
gggctggggc	cggcgggagt	ccgcgggacc	ctccagaaga	gcggccggcg	ccgtgactca	1620
gcactggggc	ggagcggggc	gggaccaccc	ttataaggct	cggaggccgc	gaggccttcg	1680
ctggagtttc	gccgccgcag	tcttcgccac	cagtgaagtac	gcgcggcccg	cgtccccggg	1740
gatggggctc	agagctccca	gcattggggc	aaccgcgacg	atcaggcccg	ggctccccgc	1800
agggtccctc	gcccacctcg	agacccgga	cgggggccta	ggggaccag	gacgtcccca	1860
gtgccgttag	cggctttcag	ggggcccgga	gcgcctcggg	gagggatggg	accccggggg	1920
cggggagggg	gggcagactg	cgctcacgcg	gccttggcat	cctcccccg	gctccagcaa	1980
acttttcttt	gttcgctgca	gtgcgcacct	acaccgtggt	ctatttccca	gttcgaggta	2040
ggagcatgtg	tctggcaggg	aaggggaggca	ggggctgggg	ctgcagccca	cagccccctg	2100
cccaccgga	gagatccgaa	cccccttacc	cctccgtcgt	gtggctttta	ccccgggcct	2160
ccttctctgt	ccccgcctct	cccgccatgc	ctgctccccg	ccccagtgtt	gtgtgaaatc	2220
ttcggaggaa	cctgtttccc	tgttccctcc	ctgcactcct	gacccctccc	cgggttgctg	2280
cgaggcggag	tgggcccggg	ccccacatct	cgtacttctc	cctccccgca	ggccgctgcg	2340
cggccctgcg	catgctgctg	gcagatcagg	gccagagctg	gaaggaggag	gtgggtgaccg	2400
tggagacgtg	gcaggagggc	tactcaaaag	cctcctgcgt	aagtgaccat	gcccggggcaa	2460
ggggaggggg	tgctgggcct	tagggggctg	tgactaggat	c		2501

<210> 2

<211> 2501

<212> DNA

<213> Homo Sapiens

<400> 2

tttgcaaatg	gagacatctt	cattattcct	atagtatcat	atgtttttta	agtttgtact	60
cacacttttg	gtgataaatg	aaggacaaga	tccttcccta	tccttgtgag	gatgactaca	120
gcatgactgg	atgggcttgc	tatgattttt	atctttccct	gtgttctcac	taccgtttta	180
ttaatctcag	ttctttttca	cagggtagca	cagaatttaa	ctagcagaaa	gagatccagc	240
catgtagacc	agagatttgt	ctaagtgcg	gcattgtaaga	atcagggaagg	aaagtttttt	300
gtttaaatac	caacaggttc	cttctctaaa	gcaattatta	tttttcaaat	ctaaccacaca	360

aggtgatagt	atccttaaac	caattaaatc	agaatctcgg	gttggataac	ctcaaataatg	420
acttattagc	acttccatt	aatcactggg	ccttcaggcc	tttaagttaa	cttactagga	480
atctcacttt	taataccatc	ttatcaactt	cagttgtaaa	taagagaaca	ctcaaaggct	540
gaggaattct	cagcggtaaa	gctctgccc	cgtaagtaa	caaaggataa	gttagtcttt	600
gttgtgatca	ctttgttgta	ctgataagct	acgtatttct	actcaaggat	tcaaattctc	660
acctttctca	agaattgggc	caaaaccgat	aaactaaact	tatttacggg	ccactgatta	720
aaggttggtg	cataataagt	tcttgctatg	ttcagcagtt	ggattcacag	cgccagaaac	780
ctataactgc	ttgactttcc	tccccactac	actgcgaaaa	ttgcccctta	aatgtaacta	840
accctaaaac	ctcaacagta	tcgtggccag	gcgtgggtgg	tcactactgt	aataccaaca	900
ttaggcatag	gcgaggggat	tgaggccagg	atatcgaaac	tagcctggga	aacacacgga	960
gacccggctc	ttggaaaaat	aattagcctt	gcgtgggtgg	gggcgcgagg	ttccggctaa	1020
tcgggaggct	acagtgcac	atgatgcac	tgcactacag	tctgcgcgac	ggcccatgtc	1080
agtaagctct	ggagcacctg	aaacaagttg	tgttgggtat	tttatttact	ggagagcgat	1140
tagtgactga	tgcctactta	cagcgactag	agacgcagtc	tccgatagca	gcacaaactc	1200
agcaggcgcg	aacaaatggg	aaagagaaac	tgggcaaaaca	agcatcacgg	ctcctcagct	1260
gagaaagtgg	gggccctaaa	aagggccttt	tgttgataga	aagggacgct	caaccaccga	1320
aaccgtagag	gggtgcggccc	tggcgcttga	gcgcgtagac	cacatccatg	gcggtgaccg	1380
tcttgcgctt	ggcgctgctt	gtataggcca	cggcgtaaccg	gatcacgttc	tccaggaaca	1440
ccttcagcac	cccgcgagtc	tcctcgtaga	tgaggccgga	gatgcgcttc	acgcgcgcgc	1500
ggcgagcaag	gcgcgggatg	gccggcttgg	tgatgccctg	gatattgtcg	cgcagtactt	1560
tacgggtggcg	cttagcgccg	cctttgccaa	gaccttccc	gcctttgccc	cgccagaca	1620
tgacgagcaa	gaggagtctc	acccaacgct	ttgtgaggac	tctggcctga	ggcagcgctt	1680
ttatagcaca	gttggcggac	cgaactgaga	acctgaaaga	agtcggcggg	aagtcccgc	1740
ccggtggggg	taaaggggcca	aaccgaaata	gggggaaaaa	aaaagcgagc	aaaagcgagc	1800
ttcttggttc	cgtgttctga	attttgtaac	gtgcatagta	ttttgttacc	acgttatgag	1860
gctttaaaaa	attgcttttg	aacgcagaag	atatacatca	atactgtggg	aaatacaaga	1920
aaggacaaga	aattaagaaa	ctacaatggt	atcccatcac	acaggctagt	taatcatgta	1980
ttttgcagag	cagttgcaca	tatttttcca	agaaaatgta	tacagtgttg	tatatggagt	2040
tttgaacctt	ccttatattg	attataattt	aaccaatttc	tattaaagag	ataaaagtga	2100
tgttttggtg	tctatgtttc	ttaggaaatta	tcaatgttta	taatcagttc	cccagcaatt	2160
ttttaatcgg	ctgtatttta	aaaataatgt	tttccacatt	caacataaat	gtactttttc	2220
tctatacttg	ggaccaatat	tgaattttat	gattttatta	caccaaattt	taaattttat	2280
tacattaata	tttaaaattg	tattagaggt	ctcatgattt	ggtactacgg	gtctccgcac	2340
tatttccctt	ccaaatttcc	taatctgttt	caccaaggtt	tctggacaac	tttagagacc	2400
ttttgtgaag	tttgaataaa	atctcttcga	gattttgata	attgcattag	ctttaggact	2460
taattggaat	agaattaaaa	tccttaaaac	aagctcttat	a		2501

<210> 3
 <211> 2251
 <212> DNA
 <213> Homo Sapiens

<400> 3

ggacccccgag	cgcgcgccag	gtagccaggga	gcggcctcag	cggcagccgc	aaactccagt	60
agccgcccgt	gctgcccgtg	gctggggcgg	agggcagcca	gagctgggga	ccaaggctcc	120
gcgccacctg	cgcgcacagc	ctcacacctg	aacgctgtcc	tcccgcagac	gagaccggcg	180
ggcactgcaa	agctgggact	cgtctttgaa	ggaaaaaaa	tagcgagtaa	gaaatccagc	240
accattcttc	actgacccat	cccgtgcac	ctcttgtttc	ccaagttttt	gaaagctggc	300
aactctgacc	tcggtgtcca	aaaatcgaca	gccactgaga	ccggctttga	gaagccgaag	360
atttggcagt	ttccagactg	agcaggacaa	ggtgaaagca	ggttggaggc	gggtccagga	420
catctgaggg	ctgaccctgg	gggtcgtga	ggctgccacc	gctgctgccc	ctacaggtag	480
gatggcggtg	ggctgacggt	ggggtcaacg	ggtagagaac	gcagggatgc	ggccctcgcc	540
gaagagagcc	aagaaggga	gagcgcgctc	tccaaattgc	ttttgtaact	tgttttcagt	600
gagcatttta	ttgatccaga	atctatcgag	aatagcacta	gcgagctact	tttcccttga	660
gatgggtcgt	attcatcttg	gcaatggagt	gagttggatt	gtggggagga	agaggaatgg	720
gaaatcagct	ttataaatat	taatgtcagc	aagagtgtgc	tgttggcagg	acgtatcgcg	780
agcctggaga	ttttgggtgg	cgcagttggg	aagtggctac	aatccagaaa	gtaggatcga	840
gttgctcccc	ttgtcttate	agtgtatcgt	ttctcggggc	cgggtctaac	accttacaag	900
tggtaatttc	cgctcacggc	agctttgtct	ctcttctacc	atcccagac	ccagccttgc	960
actccaaggc	tgccgaccgc	cagccactat	catgtccact	cccggggtca	attcgctccg	1020
ctccttgagg	cccgaaccgg	tgaacagccc	agtgaccate	ccggcggtga	tggtcatctt	1080
cgggggtggg	ggcaacctgg	tggccatcgt	gggtcgtgtg	aagtccgcga	aggagcagaa	1140
ggagacgacc	ttctacacgc	tgggtatgtg	gctggctgtc	accgacctgt	tgggcaactt	1200
gttgggtgagc	ccggtgacca	tcgccacgta	catgaagggc	caatggcccc	ggggccagcc	1260
gctgtgcgag	tacagcactt	tcattctgct	cttcttcagc	ctgtccggcc	tcagcatcat	1320
ctgcgccatg	agtgtcgagc	gctacctggc	catcaacctt	gcctattttt	acagccacta	1380
cgtggacaag	cgattggcgg	gcctcacgct	ctttgcagtc	tatgcgtcca	acgtgctctt	1440
ttgcgcgctg	cccaacatgg	gtctcggtag	ctcgcggctg	cagtaccag	acacctgggtg	1500
cttcacgcag	tggaccacca	acgtgacggc	gcacggcgcc	tactcctaca	tgtacggggg	1560
cttcagctcc	ttcctcattc	tcgccaccgt	cctctgcaac	gtgcttgtgt	gcggcgcgct	1620

gctccgcatg	caccgccagt	tcattgcgccc	cacctcgctg	ggcaccgagc	agcaccacgc	1680
ggccgcggcc	gcctcggttg	cctcccgggg	ccaccccgct	gcctccccag	ccttgccgcg	1740
cctcagcgac	tttcggcgcc	gccggagcct	ccgccgcac	gccccggccg	agatccagat	1800
ggatcatctta	ctcattgccca	cctccctggg	ggtgtcatc	tgctccatcc	cgctcgtggg	1860
gagtgaccgg	ggctggggcc	ctactcgccc	ttttctcgc	atccacctcc	cggtccatt	1920
ccccgctccc	tgcttccct	ctgagtcctt	ggcagtgaac	gtgtcgccct	taggtcgggg	1980
ctgggattcc	cacactgttt	ctcagagcag	gccccaccct	ctttgaagtc	ccaaccctaa	2040
cgagatttag	caggtgcttt	gcccctacat	ccccagttt	atgttcccgg	aagcctgggt	2100
ttctttctcc	accgagacag	cccttaccct	ttgctgcctg	acactggccg	agtcttccaa	2160
gaaaaccccc	cgccccctct	gttagacgtg	gaggggagcc	tgctgtagtg	tgacttagcc	2220
cattcctccg	tactgtgaac	tgtgaactgc	a			2251

<210> 4
 <211> 2586
 <212> DNA
 <213> Homo Sapiens

<400> 4

ccccgcgggg	cgcgggagta	gccccgctgg	gcgctcgcat	ccgcgggagt	caagccccct	60
ccccaggtgc	aggcataaaa	gtttatggct	cttgaacaat	gcggggcaga	ggtttttcca	120
agcaacgtct	aattggccgc	ttctaattaa	ggaaagagag	gcttccagct	ctatggcaac	180
ccaagcaggg	cagcttcagg	ctaaaggtag	tttagaataa	taagatcatt	ctaagaaatg	240
gaatgtctca	ctggacaccc	gaacagggtc	tctgtcattg	gaattgggtg	gtactgtact	300
tcaaccagta	ctcttgtgtg	gagggaggcg	acccagctca	ggaaagtcaa	ctacagaaag	360
aggtgacctc	cgaaaggatt	gtcttagcgc	tattagaata	catgtgacca	cacccaaaagc	420
ccaggcggac	acccgcagcc	agctcggatt	tggacaattc	aacattgctg	gcagaactga	480
agggaaacaag	ttaccccaac	cccatccctt	gtacgcgtag	tgctgagtga	ggtgggggtg	540
ggaggacagc	ggttcgttta	ttgccccctt	ttaaaatctg	agatctgaaa	atatggaggt	600
cccattcggt	ttccagctc	ttgattgccca	acaaaaaaac	aaatcccgc	ggctacattt	660
tctctctcatt	ccaaaatagc	aaccctatgg	cttgtattaa	gcccttcaga	agtttatctc	720
atgtgtctct	ggccagggag	ggaacaatgc	taggaaaagt	caccgggtgt	cttccatcct	780
cgcccccttc	aggggtgcagg	atgtgcgggc	cgggcgccct	gtgatccccg	aacgcttctc	840
gccatcccc	tgcgcgaact	tgaaggact	gggaggtgtt	gagagcagag	ttcagggtctg	900
gtgcactctg	cgggtgctgag	tgggcggcgc	gcccggcgcc	tcaggccggg	ggacctgtag	960
tgccctacc	gcgaggggga	aaatacgtag	ctggaggggc	tgccgctgc	gggttgtgat	1020
ccgttacccc	atcggtcatc	ctgggggtctc	cccaagcctc	taggtagggc	tgtgagagtc	1080
ccctagagct	gaagccccgg	aggctgacct	gtgggtctgg	ctgctatggg	aaccgggttg	1140
gtccaaagaa	gcctttcttc	cgggcacctg	gaattccagt	ttagtgtggg	gcatcgggga	1200
agtggcgctg	gggggtctgg	ttgggggacc	tcagccggca	gctccggaga	gggacctacc	1260
ttggggctgc	tgggtgaggc	cggcacgatt	cttggctcca	aaaggaaaagt	ttctgcttct	1320
tggtctggcg	cgagaagcca	aagacttatt	ttgagagcgg	agagagaaat	gttattggta	1380
acgttttctt	tggaaagttc	gagaggggtc	ttctggacac	actacctagt	gcccccaaac	1440
cagagaagta	gtttttcttt	ggtgcctggg	ctcagaagtc	gccactcact	cagcccatgg	1500
ttcgaaatca	gcatgggaag	cgccggggca	aggcttcgtc	ggagactaga	ggcctgcctg	1560
tccggaggag	cccctggggg	atggggaccc	cattctcctg	cttgcctctg	ttccacctg	1620
ggagcctcc	gtaggagccc	agaaagacga	tccactacat	ggtcccggga	cagagcagcg	1680
cgcccaactt	tgagggaact	ttgtgcgcct	ctctgaggcc	ctagctttcc	aaggcaccgc	1740
cgtccggtct	tctttcccta	gaccgaaact	ggggaagagt	gtgggcgctt	ctttgccccg	1800
atgagttcgc	ctccccaaac	gctacttctg	gctgcaccag	agcatctggg	aaactctgaa	1860
aggtgcccag	gcctcacaca	gcagcgtctc	cctactcagc	ctctgtcttt	gggttttttc	1920
aagagagctc	ctacctcatg	cctcgggtct	tcttcgatgt	cgggtccccg	aggtaggcac	1980
ggagtccctc	tgaagcagct	tgcctatctg	tgcctctttg	gtgtaaagtt	agagtttact	2040
ttgttggggg	aaggggaggt	agaaaagatc	acagttggga	aagtgcgctt	ttcgccttgt	2100
tcctaaaaca	tgcctcaaga	ctgtcatcgc	gattgttagg	agagctatca	acgtctaggg	2160
gtataaaagg	aatttctgaa	ccctcggccc	ttcccaaac	cccaggttcc	taaaacccta	2220
tcgggggtct	cttggggctg	ggattcaggc	tggcacccgt	gggaggacct	cgcttagcat	2280
ccctttatta	atatttcacg	aaggcaggct	cctgccttct	ctggagcctc	ttttctcgga	2340
atgttcccaa	actctggcta	actcactccc	ctgtgagcca	tcctagggtc	ctgtggcccc	2400
ggaagagacg	cgtcaactcc	gcgggtctgc	gcgcagtcct	tagccgcaaa	gtgctgcaag	2460
tgacccccct	gacggccctt	tccgaccgaa	gagctcggga	accaaagaga	aaaaaaataa	2520
ctttattttt	aaaagaacaa	gtcatcactg	cggcgatact	gtggcggagg	actttggcga	2580
tgggggt						2586

<210> 5
 <211> 2501
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> chemically treated genomic DNA (Homo sapiens)

<400> 5

ttgttgtata	gaatatTTTT	ttatttaggt	attatgtcga	gtatttaata	gttttttttt	60
ttgttttttt	tttttttttt	attttgtatt	ttggagttaa	ttatagtgtt	tgttgttttt	120
ttgtttgtgt	tataagtttt	tattattttag	ttttatttta	taagtggaaa	tatttagtat	180
ttggattttt	gtttttgtat	tagtttgtta	aggataatag	tttttagttt	tatttatgtt	240
tttataaaag	atatgattta	gtttttttta	atggttgtat	taaatgaagt	tttaaagata	300
taatataaat	attaattttt	tttttattat	aaaaattttt	tgttgaattt	gattataattt	360
aaattaacga	gttttgtttt	atgaaagatt	ttttggataa	atttgatagt	tgatggaata	420
ggagaagttg	tttgttatgt	ttaaagttaa	taagagatta	atatttagaa	taaatggaga	480
tttgtaaaat	aatagaaagt	aggtagtaaa	gttaaagaaa	atagttaaag	gtatagtatt	540
taaaaggaaac	gtgattatgt	tttttgtagg	gatatgggtg	gagttggaag	tcgttagttt	600
tagtaaatTT	atataggaat	agaaaattag	cgagatcgta	tggttttatt	tataagtggg	660
agttgaataa	tgagaatata	tggttatatg	gcggcgatta	atatatatgt	gtgtttgttg	720
agcgggggtg	tggggaggga	gagtattagg	aagaatagtt	aagggatatt	gggtttaata	780
tttgggtgat	gggatgattt	gtatagtaaa	ttattatggc	gtatatattt	atgtaataaa	840
tttgtatat	ttttatatgt	attttagaat	tttaaataaa	agttggacgg	ttaggcgtgg	900
tggtttacgt	ttgtaatttt	agtattttgg	gaagtcgagg	cgtgtagatt	atthaaggtt	960
aggagttcga	gattagttcg	gttaatatgg	tgaaatttcg	tttttattaa	aaatataaaa	1020
attagttaga	tgtggtacgt	atttataaatt	ttatttattc	gggaggttga	agtagaattg	1080
tttgaattcg	agaggcggag	gttgtagtga	gtcgtcgaga	tcgcgttatt	gtattttagt	1140
ttgggttata	gcgtgagatt	acgttatata	ataaaataaa	ataatataaa	ataaaaaaaa	1200
ataaaaaaaa	ataaaataaa	ataaaataaa	ataaaaaaat	aaaaataaaa	aaaaataaaa	1260
aaaataaaaat	aaagtaattt	tttttttttt	aagcggtttt	tatttttttt	ttttgttttg	1320
tgaagcgggt	gtgtaagttt	cgggatcgta	gcggtttttag	ggaatttttt	ttcgcgatgt	1380
ttcggcgcg	tagttcgttg	cgtatatatt	gttcgcgttt	tttttttggt	gtttgtttat	1440
tttttaggtt	tcgttgggga	tttgggaaag	agggaaaggt	tttttcggtt	agttgcgcgg	1500
cgatttcggg	gatttttaggg	cgtttttttg	cggtcgcagt	tcggggtgta	gcggtcgctc	1560
gggttgggg	cggcgggagt	tcgcgggatt	ttttagaaga	gcggtcggcg	tcgtgattta	1620
gtattggggc	ggagcggggc	gggattattt	ttataaggtt	cggaggtcgc	gaggttttcg	1680
ttggagtttc	gtcgtcgtag	ttttcgttat	tagtgagtac	gcgcggttcg	cgttttcggg	1740
gatgggggtt	agagttttta	gtatggggtt	aattcgtagt	attaggttcg	ggttttcgg	1800
agggtttttc	gtttatttcg	agattcggga	cgggggttta	ggggatttag	gacgttttta	1860
gtgtcggttag	cggtttttag	ggggtttcgga	gcgttttcggg	gagggatggg	atttcggggg	1920
cggggagggg	gggtagattg	cgtttatcgc	gttttgggtat	tttttttcgg	gttttagtaa	1980
attttttttt	gttcggttga	gtgtcgtttt	atatcgtggt	ttatttttta	gttcgaggtta	2040
ggagtatgtg	tttggtaggg	aagggaggtta	ggggttgggg	ttgtagttaa	tagtttttcg	2100
tttattcgga	gagattcgaa	tttttttatt	ttttcgtcgt	gtgggtttta	tttcgggttt	2160
tttttttggt	tttcggtttt	ttcgttatgt	ttgttttttcg	ttttagtgtt	gtgtgaaatt	2220
ttcggaggaa	tttggttttt	tggttttttt	ttgtattttt	gatttttttt	cgggttggtg	2280
cgaggcggag	tcgggttcgg	ttttatatatt	cgtatttttt	ttttttcgta	ggtcggttcg	2340
cggtttttcg	tatgttggtg	gtagattagg	gttagagttg	gaaggaggag	gtggtgatcg	2400
tggagacgtg	gtaggagggt	ttatttaaag	ttttttcggt	aagtgattat	gttcgggtaa	2460
ggggaggggg	tggtgggttt	taggggggtg	tgattaggat	t		2501

<210> 6

<211> 2501

<212> DNA

<213> Artificial Sequence

<220>

<223> chemically treated genomic DNA (Homo sapiens)

<400> 6

gatttttagtt	atagtttttt	aaggttttagt	attttttttt	tttgttcggg	tatggttatt	60
tacgtaggag	gtttttgagt	agtttttttg	ttacgttttt	acggttatta	tttttttttt	120
ttagtttttg	ttttgatttg	ttagtagtat	gcgtagggtc	gcgtagcgg	ttgcggggag	180
ggagaagtac	gagatgtggg	gatcgggtcg	atttcgtttc	gtagtaattc	ggggaggggg	240
taggagtgt	gggagggaat	agggaaatag	gttttttcga	agattttata	taatatggg	300
gcggggagta	ggtatggcgg	gagaggcggg	gaataggaag	gaggttcggg	gtaaaagtta	360
tacgacggag	ggataagggg	gttcggattt	tttcgggtgg	gcgaggggtt	gtgggttgta	420
gttttagttt	ttgttttttt	tttttgttag	atatatgttt	ttatttcgaa	ttgggaaata	480
gattacgggt	tagggcggta	ttgtagcgaa	ttaaagaaaag	tttggtggag	ttcgggggag	540
gatgttaagg	cgcgggtgag	gtagtttggt	tttttttttc	gttttcgggg	ttttattttt	600
tttcgaggcg	tttcgggttt	tttgaaagtc	gttaacggta	ttggggacgt	tttgggtttt	660
ttaggttttc	gttttcgggt	tcgagggtgg	cgaggagttt	tgtcgggagt	tcgggtttga	720
tggtgcgggt	tggttttatg	ttgggagttt	tgagttttat	tttcggggac	gcgggtcgcg	780
cgtattttatt	ggttcggaag	attgcggcgg	cgaattttta	gcgaagggtt	cgcgggtttc	840
gagttttata	aggggtgggt	cgttttcgtt	cgttttagtg	ttgagttacg	gcgtcggtcg	900

tttttttggg	gggtttcgcg	gatttttcgtc	ggtttttagtt	tcggcggtcg	ttgtatttcg	960
ggcgctcggtc	gtagaggggc	gtttttggagt	tttcggagtc	gtcgcgtagt	tggtcgggga	1020
agtttttttt	tttttttttag	gttttttagcg	gggttttaggg	agtaaataga	tagtaggaag	1080
aggatcgtag	cgaagtgtgc	gtagcgaatt	ggcgcgctcg	gatatcgcg	ggggaaattt	1140
tttaagatcg	ttgcgatttc	ggagtttgta	tattcgtttt	atagggtagg	ggagaggggt	1200
ggaggtcggt	tagaggaaag	gaaattgttt	tattttattt	tattttattt	tattttttta	1260
ttttatttta	ttttatttta	ttttatttta	ttttatttta	ttttatttta	ttttgtgtta	1320
ttttatttta	ttttatgacg	tagttttacg	ttgtgggtta	ggttggagtg	tagtggcgcg	1380
atctcggcgg	tttattgtaa	ttttcgtttt	tcgggtttta	gtaattttgt	tttagttttt	1440
cgagtaggtg	gaattatagg	tcggtgttat	atcttggtga	tttttgatt	tttagtagag	1500
acgggggttt	attatgttgg	tcgggttgg	ttcgaatttt	tgatttttag	tgatttgtac	1560
gtttcgggtt	tttaaagtgt	tgggattata	ggcgtgagtt	attacgtttg	gtcgtttaat	1620
ttttatttta	agttttgggg	tatatgtaga	ggatgtgtag	gtttgttata	taggtgtgtg	1680
cgttatgatg	gtttgttgta	tagattatct	tattatttag	gtattaagtt	tagtattttt	1740
tagttatttt	ttttgggtatt	tttttttttt	agtatttcgt	ttaataggta	ttagtgtgtg	1800
ttgatcgtcg	ttatgtgatt	atgtgttttt	attgttttag	ttttatttat	aagtgtgatt	1860
atgcgggttt	gtttgggtttt	gttttttctg	tgagtttctg	gaggttaacg	gttttttagtt	1920
ttattttatgt	ttttgtaaa	gatattgatta	cgtttttttt	agtggttgtg	ttttagggtta	1980
ttttttttgg	ttttgttgtt	tattttttgt	tgattttag	attttttatt	atttttagata	2040
ttgatttttt	gtttgggttta	gatattgatg	atagtttttt	ttattttatt	aattgttaag	2100
tttggttaag	gagtttttta	tgaaataaaa	ttcgttaatt	taagtgtaat	taaatttagt	2160
aagggatttt	tgtggtgggg	aagaggttgg	tgtttatgtt	gtatttttaa	aattttattt	2220
aatgtagtta	ttaaaaagaa	ttagattatg	ttttttgtgg	gaatatggat	ggagtttagag	2280
gtttattttt	ttagttaatt	aatgtaggaa	tagaaattta	aatattggat	gtttttattt	2340
gtaagtggga	gttaaatgat	gagaatttat	aatataaata	aggaaataat	agatattgtg	2400
gttgatttta	gggtgttaga	tgggaggaag	gagaggagta	gaaaagagaa	ttattgggta	2460
ttcggataaa	tatttggttg	atgaaatatt	ttgtataata	a		2501

<210> 7

<211> 2501

<212> DNA

<213> Artificial Sequence

<220>

<223> chemically treated genomic DNA (Homo sapiens)

<400> 7

tttgtaaattg	gagatatttt	tattattttt	atagtattat	atgtttttta	agtttgtatt	60
tatatatttg	gtgataaattg	aaggataaga	ttttttttta	tttttgtgag	gatgattata	120
gtatgatttg	atgggtttgt	tatgattttt	attttttttt	gtgtttttat	tatcgtttta	180
ttaaattttg	ttttttttta	tagggtagta	tagaatttaa	ttagtagaaa	gagatttagt	240
tatgtagatt	agagatttgt	ttaaagtacg	gtatgtaaga	attaggaagg	aaagtttttt	300
gtttaaatat	taatagggtt	tttttttaaa	gtaattatta	ttttttaaat	tttaatttata	360
agggtgatagt	attttttaaa	taattaaatt	agaatttcgg	gttggataat	tttaaatatg	420
atcttattagt	atctttttat	aattatttgt	tttttaggtt	tttaagttta	tttattagga	480
atcttatttt	taataatttt	ttattaattt	tagttgttaa	taagagaata	tttaagggtt	540
gaggaatttt	tagcggtaaa	gttttgttta	cgttaagtaa	taaaggataa	gttagttttt	600
gttgtgatta	ttttgttgta	ttgataagtt	acgtattttt	atcttaaggat	ttaaattttt	660
atctttttta	agaattgggt	taaaatcgat	aaattaaatt	tatttacggt	ttattgatta	720
aaggttgttg	tataataagt	ttttgttatg	tttagtagtt	ggatttatag	cgttagaaa	780
ttataattgt	ttgaattttt	tttttattat	attcgaaaaa	ttgtttttta	aatgttaatt	840
atctttaaatt	tttaatatga	tcgtgggttag	gcgtgggtgt	ttattattgt	aatattaata	900
ttaggatatg	gcgaggggat	tgaggttagg	atatacgaat	tagtttggga	aatatacggg	960
gattcgggtt	ttggaaaaat	aattagtttt	gcgtgggtgt	gggcgcgagg	tttcggttta	1020
tcggggagggt	atagttagtt	atgatgatat	tgtattatag	tttgcgcgac	ggtttatgtt	1080
agtaagtttt	ggagtatttg	aaataagttg	tgctgggtat	tttatttatt	ggagagcgat	1140
tagtgattga	tgtttattta	tagcgattag	agacgtatgt	ttcgatagta	gtataaaatt	1200
agtaggcgcg	aataaatggg	aaagagaaa	tgggtaaata	agtattacgg	tttttttagt	1260
gagaaagtgg	gggtttttaa	aaggggtttt	tggtgataga	aagggacgtt	taattatcga	1320
aatcgtagag	gggtcgggtt	tggcgtttga	gcgcgtagat	tatatttatg	gcggtgatcg	1380
ttttgcgttt	ggcgtgtttt	gtatagggtta	cggcgtttcg	gattacgttt	tttaggaata	1440
tttttagtat	ttcgcgagtt	ttttcgtaga	tgaggtcgga	gatgcgtttt	acgtcgtcgc	1500
ggcgagtaag	gcgtcgggatg	gtcgggtttg	tgatgttttg	gatattgtcg	cgtagtattt	1560
tacggtggcg	ttttgcgtcg	tttttgttaa	gatttttttc	gtttttgtcg	cggtagata	1620
tgacgagtaa	gaggagtttt	atctaacgtt	ttgtgaggat	tttggtttga	ggtagcgttt	1680
ttatacgata	gttggcggat	cgaattgaga	atctgaaaga	agtcggcggg	aagtttcgtt	1740
tcggttgggg	aggggaaatt	taaaggggtta	aatcgaaata	gggggaaaaa	aaaagcgagt	1800
tttttgtttt	cgtgttttga	atcttgtaac	gtgtatatga	ttttgttatt	acgttatgag	1860
gttttaaaaa	attgtttttg	aacgtagaag	atatatatata	atattgtggg	aaatataaga	1920
aaggataaga	aattaagaaa	ttataatgtt	atcttattat	ataggtagt	taattatgta	1980

ttttgtagag	tagttgtata	tattttttta	agaaaatgta	tatagtgttg	tatatggagt	2040
tttgtaattt	ttttatattg	attataattt	aattaatttt	tattaaagag	ataaaagtga	2100
tgttttgggtg	tttatgtttt	ttaggaatta	ttaatagtta	taattagttt	tttagtaatt	2160
ttttaatcgg	ttgtattttt	aaaataatgt	tttttatatt	taataataat	gtattttttt	2220
tttatatttg	ggattaatat	tgaaatttat	gatttttata	tattaaaatt	taaattttat	2280
tatattaata	tttaaaattg	tattagaggt	tttatgattt	ggatttacgg	gttttcgtat	2340
tatttttttt	ttaaattttt	taatttgttt	tattaaggtt	tttggataat	tttagagatt	2400
ttttgtgaag	tttgaataaa	attttttcga	gatttttgata	attgtattag	ttttaggatt	2460
taattggaat	agaattaaaa	tttttaaaat	aagtttttat	a		2501

<210> 8
 <211> 2501
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> chemically treated genomic DNA (Homo sapiens)

<400> 8

tataagagtt	tgttttaagg	attttaattt	tattttaatt	aagtttttaa	gttaatgtaa	60
ttattaaaa	ttcgaagaga	ttttatttaa	attttataaa	aggtttttaa	agttgtttag	120
aaattttggt	gaaatagatt	aggaaatttg	gaaaggaaat	aatgcggaga	ttcgtagtat	180
taaattatga	gatttttaat	ataattttta	atattaatgt	aataaaaatt	aaattttggg	240
gtaaataaat	tataaaattt	aatattgggt	ttaagtatag	agaaaaagta	tatttatggt	300
gaatgtggaa	aatattattt	ttaaaatata	gtcgattaaa	aaattgttgg	ggaattgatt	360
ataattattg	ataattttta	agaaatatag	atattaaaa	attattttta	ttttttta	420
agaaattggt	taaattataa	ttaatatag	gaggttataa	aattttatat	ataatattgt	480
atataatttt	ttggaaaaat	atgtgtaatt	gttttgtaaa	atataatgatt	aattagtttg	540
tgtagtgga	taataattgta	gtttttta	tttttgtttt	tttttgatt	ttttatagta	600
ttgatgtata	ttttttgcgt	ttaaaagtaa	ttttttaaag	ttttataacg	tggtataaaa	660
atattatgta	cgttataaaa	tttagaatac	ggaaataaga	agttcgtttt	tttttttttt	720
ttatttcggt	ttggtttttt	agattttttt	ttttttatcg	gggcgggatt	tttcgtcgat	780
tttttttagg	tttttagttc	ggttcgttaa	ttgtcgtata	aaggcgttgt	tttaggttag	840
agtttttata	aagcgttggg	tgagattttt	tttgttcgtt	atgtttgggtc	gcggtaaaag	900
cgggaagggt	tttggttaa	gcggcggtta	gcgttatcgt	aaagtattgc	gcgataatat	960
ttagggtatt	attaagtcgg	ttattcggcg	ttttgttcgt	cgcggcggcg	tgaagcgtat	1020
tttcggtttt	atttacgagg	agattcgcgg	ggtgttgaag	gtgttttttg	agaacgtgat	1080
tcgggacgtc	gtgatttata	tagagtacgt	taagcgtaa	acggttatcg	ttatggatgt	1140
ggtttacgcg	tttaagcgtt	agggtcgtat	tttttacggt	ttcgggtggt	gagcgttttt	1200
ttttattaat	aaaaggtttt	ttttaggggt	tttttttttt	tagttgagga	gtcgtgatgt	1260
ttgtttgttt	agtttttttt	tattatttgt	tcgcgtttgt	tgagtttgtg	ttgttatcgg	1320
agtatgcgtt	tttagtcgtt	gtaagtaggt	attagttatt	aatcgttttt	tagtaaataa	1380
aatatttta	ataatttggt	ttaggtgttt	tagagtttat	tgatatgggt	cgtcgcgtag	1440
attgtagtgt	agtgttatta	tggtttattg	tagtttttcg	attagtcgga	atttcgcgtt	1500
tattattacg	taaggtta	tattttttta	aagatcgggt	tttcgtgtgt	tttttaggtt	1560
agtttcgata	ttttgggttt	aatttttttcg	tttatgttta	atgttggtat	tatagtagtg	1620
agttattacg	tttggttacg	atattgttga	ggtttttaggg	ttagttatat	tttaaggggt	1680
attttcgtag	tgtagtgagg	aggaaagtta	agtagttata	ggtttttggc	gttggtgaatt	1740
taattgttga	atatagtaa	aattttattat	gtaataaatt	tttaattagtg	gatcgtaaat	1800
aagtttagtt	tatcggtttt	ggtttaattt	ttgagaaagg	tgagaatttg	aatttttgag	1860
tagaaatacg	tagtttatta	gtataataaa	gtgattataa	ttaaagattaa	tttatttttt	1920
gttatttaac	gtgggtagag	ttttatcgtt	gagaattttt	tagtttttga	gtgttttttt	1980
atttataatt	gaagttgata	agatgggtatt	aaaagtgaga	tttttagtaa	gtaaatttaa	2040
aggtttgaag	gattagtgat	taatgggaag	tgtaataaag	ttatatttga	ggttatttaa	2100
ttcgagattt	tgatttaatt	ggtttaagga	tattattatt	ttgtgggtta	gatttgaaaa	2160
ataataattg	ttttaaggaa	ggaatttggt	ggatttttaa	taaaaaattt	tttttttgta	2220
tttttatatg	tcgttattta	gataaatttt	tggtttatat	ggtttgattt	tttttttgta	2280
gttaaatttt	gtgttatttt	gtgaaaaaga	attgagatta	ataaacgggt	agtgagaata	2340
taggggaaaga	taaaaattat	agtaagttta	tttagttatg	ttgtagttat	ttttataagg	2400
ataggggaagg	attttgtttt	ttatttatta	tttaaggtgt	gagtataaat	tttaaaaaata	2460
tatgatatta	taggaataat	gaagatgttt	ttatttgtaa	a		2501

<210> 9
 <211> 2251
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> chemically treated genomic DNA (Homo sapiens)

<400> 9

ggatttcgag	tcggttttag	gtagtttagga	gcggttttag	cggtagtcgt	aaatttttagt	60
agtcggttcgt	gttggttcgtg	gttggggcgg	agggtagtta	gagttgggga	ttaaggtttc	120
gcggtattttg	cgcgtatagt	tttatatttg	aacgttggtt	tttcgtagac	gagatcggcg	180
ggtattgtaa	agttgggatt	cgtttttgaa	ggaaaaaaa	tagcgagtaa	gaaatttagt	240
attattttttt	attgatttat	ttcgttggtat	tttttggttt	ttaagttttt	gaaagttggt	300
aattttgatt	tcggtgttta	aaaatcgata	gttattgaga	tcggttttga	gaagtcgaag	360
atttggtagt	ttttagattg	agtaggataa	ggtgaaagta	ggttggaggc	gggttttagga	420
tatttgaggg	ttgatttttg	gggttcgtga	ggttggtatc	gttggtgtcg	ttataggtga	480
gatggcggtg	ggttgacgtt	gggggttaacg	ggtagagaac	gtagggatgc	ggtttttcgtc	540
gaagagagtt	aagaagggaa	gagcgcgttt	tttaaattgt	ttttgtaatt	tggttttagt	600
gagtatttta	ttgatttaga	atttatcgag	aatagtatta	gcgagttatt	ttttttttga	660
gatgggtttt	atttattttg	gtaatggagt	gagttggatt	gtggggagga	agaggaatgg	720
gaaaattagt	ttataaatat	taatgttagt	aagagtgtgt	tggtggtagg	acgtatcgcg	780
agtttgagga	ttttgggtgt	cgtagttggt	aagtggttat	aatttagaaa	gtaggatcga	840
gttggttttt	ttgttttatt	agtgtatcgt	ttttcgggcg	cgggtttaat	attttataag	900
tggtaaattt	cgtttacggt	agttttgttt	tttttttatt	attttttagat	ttagttttgt	960
attttaaggt	tgcttatcgt	tagttattat	tatgtttatt	ttcgggggta	attcgttcgt	1020
ttttttgagt	ttcgatcggg	tgaatagttt	agtgtattat	tcggcggtga	tggtttatttt	1080
cggggtggtg	ggttaatttg	tggttatcgt	ggtgttggtg	aagtcgcgta	aggagtagaa	1140
ggagacgatt	ttttatacgt	tggtatgtgg	gttggttggt	atcgatttgt	tggttatttt	1200
gttggtgagt	tcggtgatta	tcgttacgta	tatgaagggt	taatggttcg	ggggttagtc	1260
gttggtcgag	tatagtattt	ttattttggt	tttttttagt	ttgttcgggt	ttagtattat	1320
ttgcgttatg	agtgtcgagc	gttatttggt	tattaattat	gtttattttt	atagttatta	1380
cgtggataag	cgattggcgg	gttttacggt	ttttgtagtt	tatgcgttta	acgtgttttt	1440
ttgcgcgttg	tttaaatatg	gtttcggtag	ttcgcgggtg	tagtatttag	atatttggtg	1500
ttttatcgat	tggattatta	acgtgacggc	gtacgtcgtt	tatttttata	tgtagcgggg	1560
tttttagtttt	ttttttattt	tcgttatcgt	tttttgtaac	gtgtttgtgt	gcggcgcgtt	1620
gtttcgtagt	tatcgtagt	ttatgcgtcg	tatttcggtg	ggtagcgagt	agtattacgc	1680
ggtcgcgggtc	gtttcgggtg	tttttcgggg	ttatttcggt	gttttttttag	ttttgtcgcg	1740
tttttagcgat	tttcggcggtc	gtcggagttt	tcgtcgatc	gcgggcgtcg	agatttagat	1800
ggttattttta	tttattgtta	tttttttggt	ggtgtttatt	tggtttattt	cgttcgtggt	1860
gagtgatcgg	ggttgggggt	ttattcgggt	tttttttcgt	atttattttt	cgcggttatt	1920
tttcggtttt	tggttttttt	ttgagttttt	ggtagtgaac	gtgtcggttt	taggtcgggg	1980
ttgggatttt	tatatgtttt	tttagagtag	gttttaatttt	ttttgaagtt	tttaattttta	2040
cgagatttag	taggtgtttt	gtttttatat	tttttagttt	atgttttcgg	aagtttgggt	2100
tttttttttt	atcgagatag	tttttatttt	ttgttggttg	atattgggtcg	agtttttttaa	2160
gaaaattttt	cgtttttttt	gttagacgtg	gaggggagtt	tggtgtagtg	tgatttagtt	2220
tatttttttcg	tattgtgaat	tgtgaattgt	a			2251

<210> 10

<211> 2251

<212> DNA

<213> Artificial Sequence

<220>

<223> chemically treated genomic DNA (Homo sapiens)

<400> 10

tgtagtttat	agtttatagt	acggaggaat	gggttaagtt	atattatagt	aggttttttt	60
ttacgttttaa	tagagggggc	gggggggttt	tttgaagat	tcggttagtg	ttaggtagta	120
aggggtaagg	gttggttcgg	tgagagaaaga	aatttaggtt	ttcgggaata	ttaaattgggg	180
gatgtagggg	taaagtattt	gttaaatttc	gttaggggtg	ggattttaaa	gaggggtggg	240
tttggttttga	gaaatagtgt	gggaatttta	gtttcgattt	aaaggcgata	cgtttattgt	300
taaggattta	gagggaaagt	agggagcggg	gaatggacgc	gggaggtgga	tgcgagaaaa	360
aggtcgagta	gggttttagt	ttcggttatt	tattacgagc	gggatggagt	agatgagtat	420
tattagggag	gtggtaatga	gtaagatgat	tatttggtat	tcggcggttcg	cgatgcggcg	480
gaagtttcgg	cggcgctcga	agtcgttag	gcgcggtga	gttggggagg	tagcgggggtg	540
gtttcgggag	gtaatcgagg	cggtcgcggt	cgcgtgggtg	tggttcggtg	ttagcgagggt	600
gcggcgtagt	aattggcggt	gtatgcggag	tagcgcgtcg	tatataagta	cgttgtagag	660
gacgggtggcg	agaatgagga	aggagttaga	gttcgcgtat	atgtaggagt	aggcggcggtg	720
cgtcggttacg	ttggtggttt	agtcgatgaa	gtattagggt	tttgggtatt	gtagtcgcga	780
gttatcgaga	tttatgttgg	gtagcgcgta	aaagagtacg	ttggacgtat	agattgtaaa	840
gagcgtgagg	ttcgttaatc	gtttgtttac	gtagtgggtg	tagaaaatagg	tatgggtgat	900
ggttaggtag	cgttcgatat	ttatggcgta	gatgatgttg	aggtcggata	ggttgaagaa	960
gagtagaatg	aagggtgtgt	attcgtatag	cggttgggtt	tcgggttatt	ggttttttat	1020
gtacgtggcg	atgggtatcg	ggtttattaa	taaagtgttt	aataggtcgg	tgatagttag	1080
tttatatatt	agcgtgtaga	aggcgtgttt	ttttgttttt	ttgcgcgatt	tgtatagtat	1140
tacgatgggt	attaggttgt	ttattatttc	gaagatgaat	attatcgctcg	ggatgggtat	1200

tgggttggtt	agtcggtcgg	ggtttaagga	ggcggacgaa	ttgatttcgg	gagtggatat	1260
gatagtggtt	ggcgggtcgt	agttttggag	tgtaagggtg	ggtttgggga	tggtagaaga	1320
gagataaagt	tgctgtgagc	ggaaattatt	at ttgt aagg	tgtagattc	gcgttcgaga	1380
aacgataat	tgataagata	aggggagtaa	ttcgatttta	ttttttggat	tgtagttatt	1440
tatttaattgc	ggttattaaa	at tttt taggt	tcgcgatacg	ttttgttaat	agtataattt	1500
tggtgatatt	aataattata	aattgatttt	tttatttttt	ttttttttta	taatttaatt	1560
tattttattg	ttaagatgaa	taagatttat	tttaagggaa	aagtagttcg	ttagtgttat	1620
tttcgataga	ttttgaatta	ataaaatggt	tattgaaaat	aagttataaa	agtaatttgg	1680
agagcgcgtt	tttttttttt	tgggtttttt	cggcgagggt	cgtatttttg	cgttttttat	1740
tcgttgattt	taacgttagt	ttaacgttat	tttatttgta	gcggtagtag	cggtggtagt	1800
tttacgagtt	tttaggggta	gttttttagt	gtttttggat	cgtttttaat	ttgtttttat	1860
tttgttttgt	ttagtttgga	aattgttaaa	ttttcgggtt	tttaaagtcg	gttttagtgg	1920
ttgtcgattt	ttgggatacg	aggttagagt	tgtaggtttt	taaaaatttg	ggaaataaga	1980
gggttagcgg	gatggggttag	tgaagaatgg	tggtggattt	tttattcgtt	at tttttttt	2040
ttttaaagac	gagtttttagt	ttttagtgtt	tcgtcgggtt	cgtttgcggt	aggatagcgt	2100
ttaggtgtga	ggttgtgcgc	gtaggtggcg	cggaggtttt	gttttttagt	ttggttggtt	2160
ttcgttttag	ttacgggttag	tacgggcggt	tattggagtt	tcgggttgtc	gttgaggctg	2220
tttttggtta	tttgggggcg	gttcgggggt	t			2251

<210> 11

<211> 2586

<212> DNA

<213> Artificial Sequence

<220>

<223> chemically treated genomic DNA (Homo sapiens)

<400> 11

ttcgcgcggg	cgcgggagta	gtttcgttgg	gcgttcgtag	tcgcgggagt	taagtttttt	60
ttttaggtgt	aggtataaaa	gtttatggtt	tttgaataat	gcggggtaga	ggttttttta	120
agtaacgttt	aattggctcg	ttttaattaa	ggaaagagag	gttttttagt	ttatggtaat	180
ttaagtaggg	tagtttttag	ttaaagggtat	tttagaataa	taagattatt	ttaagaaatg	240
gaatgtttta	ttggatattc	gaataggttt	tttgttattg	gaattgggtg	gtattgtatt	300
ttaattagta	tttttgtgtg	gagggaggcg	at tttagttt	ggaaagttaa	ttatagaaag	360
aggtgatttt	cgaaggattt	gttttagcgt	tattagaata	tatgtgatta	tattaaaagt	420
ttaggcggat	attcgtagtt	agttcggatt	tggataattt	aataattgtt	gtagaattga	480
aggggaataa	ttattttta	tttatttttt	gtacgcgtag	tggttagtga	gttgggggtg	540
ggaggatagc	ggttcgttta	ttgttttttt	ttaaaatttg	agatttgaaa	atatggagggt	600
tttatcgtt	tttttagttt	ttgattgtta	ataaaaaaat	aaatttcgtt	ggttatatatt	660
tttttttatt	ttaaaatagt	aattttatgg	tttgtattaa	gtttttttaga	agttttattt	720
at ttgttttt	ggtttagggag	ggaataatgt	taggaaaagt	tatcgggtgt	tttttatttt	780
cgtttttttt	aggggtgtag	atgtgcgggt	cggcgggttt	gtgatttcgg	aacgtttttt	840
gttatttttt	tcgcgcgaatt	tgaaggattt	gggaggtgtt	gagagtagag	tttaggggtg	900
gtgtattttg	cgggtgttag	tgggcggcgc	gttcgggcgt	ttaggtcggg	ggattttagt	960
tcgtttttat	cgcggagggga	aaatacgtag	ttggagggcg	tcggtcgtgc	gggttgtagt	1020
tcgttatttt	atcgtttatt	ttgggggttt	tttaagtttt	taggttaggt	tgtagagatt	1080
tttttagagt	gaagtttcgg	aggttgattt	gtgggttttg	ttgttatggg	aattcgggtg	1140
gtttaaagaa	gttttttttt	cgggtatttg	gaatttttag	ttagtgtggg	gtatcggggg	1200
agtggcgttg	gggggttggg	ttgggggatt	ttagtccgta	gtttcggaga	gggtttattt	1260
ttgggggtcgt	tgggttaggt	cgggtacgatt	tttggtttta	aaaggaaaagt	ttttgttttt	1320
tgttttggcg	cgagaagtta	aagattttatt	tttagagcgg	agagagaaat	gttattggta	1380
acgttttttt	tgaaaagttc	gagaggggtt	ttttggatat	attatttagt	gtttttaaat	1440
tagagaagta	gttttttttt	ggtgttttgg	tttagaagtc	gttatttatt	tagtttatgg	1500
ttcgaaatta	gtatgggaag	cgtcggggta	aggtttcgtc	ggagattaga	ggtttgtttg	1560
tcgggaggag	tttttggggg	atggggattt	tatttttttg	tttgtttttg	tttttatttg	1620
ggacgttttc	gtaggagttt	agaaagacga	tttattatat	ggtttcggga	tagagttagc	1680
cgtttaattt	tgagggaatt	ttgtgcgttt	tttttaggtt	ttagtttttt	aagggtatcgt	1740
cgttcgtttt	ttttttttta	gatacgaaatt	ggggaagagt	gtgggcgttt	ttttgtttcg	1800
atgagttcgt	ttttttaaac	gtttattttc	gttggtattag	agtatttggg	aaattttgaa	1860
aggtgttttag	gttttatata	gtagcgtttt	tttatttagt	ttttgttttt	gggttttttt	1920
aagagagttt	ttattttatg	tttcgggttt	ttttcgtatg	cgggttttcg	aggttaggtac	1980
ggagtttttt	tgaaaagtagt	tgtttatttg	tgtttttttg	gtgtaaagtt	agagtttatt	2040
ttgttggggg	aaggggaggt	agaaaagatt	atagttggga	aagtgcgttt	ttcgttttgt	2100
ttttaaaata	tgttttaaga	ttgttatcgc	gattgttagg	agagttatta	acgttttaggg	2160
gttataaagg	aatttttgaa	ttttcgggtt	tttttaaat	tttaggtttt	taaaatttta	2220
gtgggggttt	tttgggggtg	ggatttaggt	tgggtatcgt	gggaggattt	cgtttagtat	2280
ttttttatta	atattttacg	aaggtaggtt	tttggttttt	ttggagtttt	ttttttcggg	2340
atgttttttaa	at tttaggtta	at ttattttt	ttgttagtta	ttttaggttt	ttgtgggttcg	2400
ggaagagacg	cgttaatttc	gcgggtttgc	cgtgagtttt	tagtcgtaaa	gtgttgtaag	2460
tgattttttt	gacgggtttt	ttcgatcgaa	gagttcggga	attaaagaga	aaaaaaataa	2520

ttttatTTTT aaaagaataa gttattattg cggcgatatt gtggcggagg attttggcga	2580
tggggt	2586

<210> 12
 <211> 2586
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> chemically treated genomic DNA (Homo sapiens)

<400> 12

atTTTatcgt	taaagTTTT	cgTtatagta	tcgtcgtagt	gatgatttgt	TTTTttgaaa	60
ataaagttat	TTTTTTTTTT	ttTggTTTTc	gagTTTTtcg	gtcggaaagg	gtcgttaggg	120
gggttatttg	tagtattttg	cggTTaaagga	ttgcgcgtag	attcgcggag	ttgacgcgtt	180
TTTTttcggg	ttatagagtt	ttaggaTggT	ttatagggga	gtgagttagt	tagagTTtgg	240
gaatatttTc	agaaaagagg	TTTTtagagaa	ggtaggagtt	Tgttttcgtg	aaatattaat	300
aaagggatgt	taggcgaggt	TTTTttagcg	gtgttagttt	gaattttagt	tttaagagat	360
TTTTattagg	gttttaggaa	ttTgggggtt	tgggaagggt	cgagggTTta	gaaattTTTT	420
tatagTTTTt	agacgttgat	agTTTTTTta	ataatcgCga	TgatagTTTT	gaggTatgtt	480
ttaggaataa	ggcgaaaagc	gtattTTTTt	aattgtgatt	TTTTttattt	TTTTTTTTTT	540
taataaagta	aattTTaatt	ttatattaaa	gggtataga	taggtAattg	TTTTtagagg	600
gaattcgtgt	ttatttcggg	gattcgatat	cgaagaaaga	tcgaggTatg	aggtagagat	660
TTTTttgaaa	aaatttaaag	atagaggttg	agtagggaga	cgTtgtTgtg	TgaggTTtgg	720
gtattTTTTa	gagTTTTTTa	gatgtTTTTg	TgtagtCgaa	gtaggcgTTt	ggggaggCga	780
atTTtatcgg	gtaaagaagc	gtttatattt	TTTTttagtt	tcggtTTtagg	gaaagaagaa	840
cggacggcgg	Tgttttgga	agttagggTt	ttagagaggc	gtataaagtt	TTTTtaaagt	900
Tgggcgcgtt	gttttgTttc	gggattatgt	agtggatcgt	TTTTttgggt	TTTTacggag	960
gcgtttttag	Tgggaattag	agtaagtagg	agaatggggt	TTTTattttt	taggggtTTt	1020
tttcgatagg	taggtTTTTa	gttttcgacg	aagttttgtt	tcggcgTTtt	ttatgttgat	1080
ttcgaattat	gggttgagtg	agtggcgatt	tttgagTTta	ggtattaaag	aaaaattatt	1140
TTTTtggttt	gggggtatta	ggtagtgtgt	ttagaagatt	TTTTtcgaat	TTTTtaaaga	1200
aaacgtttat	aataatattt	TTTTtttcgt	TTTTaaaata	agTTTTtggt	TTTTcgcgTt	1260
agaataagaa	gtagaaattt	TTTTtttgga	gttaagaatc	gtgtcggttt	tatttagCga	1320
TTTTaaagggt	aggTTTTTTt	cggagttgtc	ggTtgaggTt	TTTTaattta	gttttttagc	1380
gttattTTTT	cgatgtTTTa	tattaaattg	gaattttagg	Tgttcggaag	aaaggTTTTt	1440
ttggattaat	cgggtTTTTa	tagtagttag	atTTtataggt	tagTTTTtcgg	ggTTTTagtt	1500
ttaggggatt	tttatagttt	tatttagagg	tttggggaga	TTTTaggatg	atcgatgggg	1560
taacggatta	taattcgtac	ggcgtacgtt	TTTTagttac	gtattTTTTt	tttcgcggta	1620
gggcgattat	aggtTTTTtc	gtttgagcgt	tcgggcgcgt	cgTTTTattta	gtatcgtaga	1680
gtgtattagt	tttgaatttt	gtttttaata	TTTTtttagt	TTTTtaagtt	cgcgtaaggg	1740
gatggttagga	agcgttttcg	gattataggt	tcgtcggTtc	gtatattTTt	tattttggaa	1800
ggggcgagga	Tggaagagta	tcggtgattt	TTTTtagtat	TgtTTTTTTt	ttggtttaga	1860
gtaaatgaga	taaattTTTT	aagggtTTta	tataagttat	agggttgTta	TTTTggaatg	1920
aggagaaaat	gtagttagcg	ggatttgTtt	TTTTgttggt	aattaaagat	Tgggaaaacg	1980
aatgggattt	TtatattTTt	agattTTtaga	TTTTaaaagg	gggtAataaa	cgaatcgTtg	2040
TTTTTTTTatt	TTtaattttat	ttagtattac	gcgtataggg	gatggggTtg	gggtAatttg	2100
TTTTTTTTtag	TTTTgttagt	aatgttgaat	Tgtttaaatt	cgagttgggt	gcgggtgttc	2160
gtttgggttt	ttggtgtggT	tatatgtatt	TTaatagcgt	taagataatt	TTTTcgaggg	2220
ttattTTTTt	ttgtagtTga	TTTTTTTTaga	ttgggtcgTt	TTTTTTTTata	taagagtatt	2280
ggttgaagta	tagtatatat	taattTTaat	gatagagaat	ttgttcgggt	gttttagTgag	2340
atattttatt	TTTTagaatg	atTTttattat	TTtaaagtat	TTTTagtttg	aagtTgtttt	2400
gtttgggttg	ttatagagtt	ggaagTTTTt	TTTTTTTTtaa	ttagaagcgg	TTaattagac	2460
gttgTttTgga	aaaattTTTT	tttcgtattg	TTtaagagtt	ataaattTTt	atgtttgtat	2520
ttggggaggg	ggTTtgattt	tcgcggTtgc	gagcgtTTtag	cggggttatt	ttcgcgttcg	2580
cgcggg						2586

<210> 13
 <211> 2501
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> chemically treated genomic DNA (Homo sapiens)

<400> 13

ttgttgTata	gaatattTTTa	ttattttaggt	attatgtTga	gtattttaata	gtttttTTTT	60
ttgtttTTTT	TTTTTTTTTT	atTTttgtatt	ttggagTTaa	ttatagtTgt	TgttgTTTTt	120
ttgtttgtgt	tataagTTTT	tattattttag	TTTTattTTa	taagtGagaa	tatttagTat	180

ttggattttt	gtttttgtat	tagtttggtta	aggataatag	tttttagttt	tatttatgtt	240
tttataaaag	atatgattta	gtttttttta	atgggtgtat	taaatgaagt	tttaaagata	300
taatataaat	attaattttt	tttttattat	aaaaattttt	tgttgaattt	gatttatatt	360
aaattaatga	gtttttgttt	atgaaagatt	ttttggataa	atttgatagt	tgatggaata	420
ggagaagtgg	tttgttatgt	ttaaagttaa	taagagatta	atatttagaa	taaattggaga	480
tttgtaaatt	aatagaaagt	aggtagtaaa	gttaaagaaa	atagtttaag	gtatagttat	540
taaaaggaat	gtgattatgt	tttttgtagg	gatatgggtg	gagttggaag	ttgttagttt	600
tagtaaat	atataggaat	agaaaattag	tgagattgta	tggttttatt	tataagtggg	660
agttgaataa	tgagaatata	tggttatatg	gtgggtgatta	atatatat	gtgtttgttg	720
agtggggtgt	tggggaggga	gagtattagg	aagaatagtt	aagggatatt	gggtttaata	780
tttgggtgat	gggtagattt	gtatagtaaa	ttattatggt	gtatatattt	atgtaataaa	840
tttgtatatt	ttttatatgt	attttagaat	tttaaataaa	agttggatgg	ttaggtgtgg	900
tggtttatgt	ttgtaatttt	agtatttttg	gaagttgagg	tgtgtagatt	atttaaggtt	960
aggagtttga	gatttagttt	gttaatatgg	tgaatttttg	tttttattaa	aaatataaaa	1020
attagttaga	tgtggtatgt	atttataatt	ttatttattt	gggaggttga	agtagaattg	1080
tttgaatttg	agaggtggag	gttgtagtga	gttgttgaga	ttgtgttatt	gtatttttagt	1140
ttgggttata	gtgtgagatt	atgttataaa	ataaaataaa	ataatataaa	ataaaataaa	1200
ataaaataaa	ataaaataaa	ataaaataaa	ataaaataaa	ataaaaaaat	aaaaataaat	1260
aaaaataaat	aaagtaattt	tttttttttt	aagtggtttt	tatttttttt	ttttgttttg	1320
tgaagtgggt	gtgtaagttt	tgggatttga	gtggtttttag	ggaatttttt	tttgtgtatg	1380
tttgggtgtg	tagtttgttg	tgtatatatt	gttgtgggtt	tttttttgtt	gtttgtttat	1440
tttttaggtt	ttgttgggga	tttgggaaag	aggggaaaggt	tttttttgggt	agttgtgttg	1500
tgattttggg	gatttttaggg	tgtttttttg	tggttgatgt	ttggggtgta	gtgggtgttg	1560
gggttggggg	tgggtgggagt	ttgtgggatt	ttttagaaga	gtgggtgggtg	ttgtgattta	1620
gtattggggg	ggagtggggg	gggattattt	ttataaggtt	tggaggttgt	gaggtttttg	1680
ttggagtttt	gttgtttgtag	tttttgttat	tagtgagtat	gtgtgggttg	tgtttttggg	1740
gatgggggtt	agagttttta	gtatgggggt	aatttgtagt	attaggtttg	ggtttttgggt	1800
agggtttttt	gtttattttg	agatttggga	tgggggttta	ggggatttag	gatgttttta	1860
gtgttgttag	tggtttttag	ggggtttggga	gtgttttggg	gagggatggg	attttggggg	1920
tggggagggg	gggtagattg	tgtttattgt	gttttgggat	tttttttggg	gttttagtaa	1980
attttttttt	gtttgttgta	gtgttgtttt	atattgtggt	ttatttttta	gtttgaggta	2040
ggagtatgtg	tttggtaggg	aaggagggtg	gggggttggg	ttgtagttta	tagttttttg	2100
tttatttggg	gagatttgaa	tttttttatt	tttttgttgt	gtgggtttta	ttttgggttt	2160
tttttttgtt	ttttgttttt	tttgttatgt	ttgttttttg	ttttagtgtt	gtgtgaaatt	2220
tttggaggaa	tttgtttttt	tgtttttttt	ttgtattttt	gatttttttt	tgggttgttg	2280
tgaggtggag	ttgggttggg	ttttatat	tgtatttttt	tttttttgtg	gggttgtgtg	2340
tgggtttgtg	tatgttgttg	gtagattagg	gttagagttg	gaaggaggag	gtgggtgattg	2400
tggagatgtg	gtaggagggt	ttatttaaag	ttttttgtgt	aagtgattat	gtttgggtaa	2460
ggggaggggg	tgttgggttt	taggggggtg	tgattaggat	t		2501

<210> 14

<211> 2501

<212> DNA

<213> Artificial Sequence

<220>

<223> chemically treated genomic DNA (Homo sapiens)

<400> 14

gatttttagtt	atagtttttt	aagggttagt	attttttttt	tttgtttggg	tatggttatt	60
tatgtaggag	gttttgagt	agtttttttg	ttatgttttt	atggttatta	tttttttttt	120
ttagttttgg	ttttgatttg	ttagtagtat	gtgtagggtt	gtgtagtggg	ttgtggggag	180
ggagaagtat	gagatgtggg	gattgggttg	attttgtttt	gtagtaattt	ggggaggggt	240
taggagtgtg	gggagggaat	agggaaatag	gtttttttga	agattttata	taataattgg	300
gtggggagta	ggtatggtgg	gagaggtggg	gaataggaag	gaggtttggg	gtaaaagtta	360
tatgatggag	ggataagggg	gtttggattt	ttttgggtgg	gtgaggggtt	gtgggttgta	420
gttttagttt	ttgttttttt	tttttgttag	atatatgttt	ttattttgaa	ttgggaaata	480
gattatggtg	tagggtggta	ttgtagtga	taaagaaaag	tttgttggag	tttgggggag	540
gatgttaagg	tgtggtgagt	gtagtttgtt	tttttttttt	gttttttggg	ttttattttt	600
ttttgaggtg	ttttgggttt	tttgaaagtt	gttaatggta	ttggggatgt	tttgggtttt	660
ttaggttttt	gttttgggtt	ttgaggtggg	tgaggagttt	tgttgggagt	ttgggtttga	720
tgttgtgggt	tgggttttatg	ttgggagttt	tgagttttat	ttttggggat	gtgggttgtg	780
tgtattttatt	ggtggtgaag	attgtggtgg	tgaattttta	gtgaagggtt	tgtgggtttt	840
gagttttata	agggtgggtt	tgttttgttt	tgttttagtg	ttgagttatg	gtgttgggtg	900
tttttttggg	gggttttgtg	gatttttgtt	ggtttttagt	ttggtgggtg	ttgtattttg	960
ggtgttgggt	gtagaggggt	gttttggagt	ttttggagtt	gttgtgtagt	tgggtgggga	1020
agtttttttt	tttttttttag	gttttttagt	gggttttagg	agtaaataga	tagtaggaag	1080
aggattgtag	tgaagtgtgt	gtagtgaatt	gggtgtgtgg	gatattgtgg	ggggaaattt	1140
tttaagattg	ttgtgatatt	ggagtttgta	tatttgtttt	atagggtagg	ggagaggggt	1200
ggaggttgtt	tagaggaaa	gaaattgttt	tattttattt	tattttattt	tattttttta	1260

ttttatttta	ttttatttta	ttttatttta	ttttatttta	ttttatttta	ttttgtgtta	1320
ttttatttta	ttttatgatg	tagttttatg	ttgtgggtta	ggttggagtg	tagtgggtgtg	1380
attttgggtg	tttattgtaa	tttttgtttt	ttgggtttaa	gtaattttgt	tttagttttt	1440
tgagtaggtg	gaattatagg	tgtgtgttat	atttgggtga	tttttgtatt	tttagtagag	1500
atgggggttt	attatgttgg	ttgggttggg	tttgaatttt	tgattttagg	tgattttgat	1560
gttttgggtt	tttaaagtgt	tgggattata	gggtgtgagtt	attatgtttg	gttgtttaat	1620
ttttatttga	agttttgggg	tatatgtaga	ggatgtgtag	gtttgttata	taggtgtgtg	1680
tgttatgatg	gtttgttgta	tagattattt	tattattttag	gtattaagtt	tagtattttt	1740
tagttatttt	ttttgggtatt	tttttttttt	agtattttgt	ttaataggta	ttagtgtgtg	1800
ttgattgttg	ttatgtgatt	atgtgttttt	attgtttagt	ttttatttat	aagtgtgatt	1860
atgtggtttt	gttgggtttt	tgtttttgtg	tgagtttggt	gaggttaatg	gttttttagtt	1920
ttatttatgt	ttttgtaaag	gatatgatta	tgtttttttt	agtgggtgtg	ttttaggtta	1980
tttttttttg	ttttgttggt	tattttttgt	tgattttag	attttttatt	attttagata	2040
ttgatttttt	gttgggttta	gatatgatag	atagtttttt	ttattttatt	aattgttaag	2100
tttgttttaag	gagtttttta	tgaataaaaa	tttgttaatt	taagtgtaat	taaatttagt	2160
aagggatttt	tgtgggtggg	aagagggttg	tgtttatggt	gtatttttaa	aatttttagt	2220
aatgtagtta	ttaaaaagaa	ttagattatg	tttttgggtg	gaatatggat	ggagtttagg	2280
gttattattt	ttagttaaatt	aatgtaggaa	tagaaaattt	aatattggat	gtttttattt	2340
gtaagtggga	gttaaattgat	gagaatttat	aatataaata	aggaaataat	agatattgtg	2400
gttgatttta	gggtgtagga	tgggaggaag	gagaggagta	gaaaagagaa	ttattgggta	2460
tttgggtata	tatttgggtg	atgaaatatt	ttgtataata	a		2501

<210> 15

<211> 2501

<212> DNA

<213> Artificial Sequence

<220>

<223> chemically treated genomic DNA (Homo sapiens)

<400> 15

tttgtaaatg	gagatatattt	tattattttt	atagtattat	atgtttttta	agtttgtatt	60
tatatatttg	gtgataaatg	aaggataaga	ttttttttta	tttttgtgag	gatgattata	120
gtatgattgg	atgggtttgt	tatgattttt	attttttttt	gtgtttttat	tattgtttta	180
ttaattttag	ttttttttta	tagggtagta	tagaatttaa	ttagtagaaa	gagatttagt	240
tatgtagatt	agagatttgt	ttaatgtatg	gtatgtaaag	attaggaagg	aaagtttttt	300
gtttaaatat	taatagggtt	tttttttaaa	gtaattatta	ttttttaaat	ttaatttata	360
aggtgatagt	atttttaaat	taattaaatt	agaatttttg	gttggataat	tttaaatatg	420
attttatagt	attttttatt	aattattggt	tttttaggtt	tttaagttta	tttattagga	480
attttatttt	taataattatt	ttattaattt	tagttgtaaa	taagagaata	tttaaagggt	540
gaggaatttt	tagtgggtaaa	gttttgttta	tgtaaagtaa	ttaaaggata	gttagttttt	600
gttgtgatta	ttttgttgta	ttgataagtt	atgtattttt	atttaaggat	ttaaattttt	660
atttttttta	agaattgggt	taaaattgat	aaattaaatt	tatttatggt	ttattgatta	720
aaggttggtg	tataataagt	ttttgttatg	tttagtagtt	ggatttatag	tgtagaaaat	780
ttataattgt	ttgatttttt	ttttttattt	attgtgaaaa	ttgtttttta	aatgtaatta	840
attttaaaaa	tttaatatga	ttgtgggttag	gtgtgggtgg	ttattattgt	aatattaata	900
ttaggtatag	gtgaggggat	tgaggttagg	atattgaaat	tagtttggga	aatatatgga	960
gatttgggtt	ttggaaaaat	aattagtttt	gtgtgggtgg	gggtgtgagg	ttttgggtta	1020
ttgggagggt	atagtgagtt	atgatgatat	tgtattatag	tttgtgtgat	ggtttatggt	1080
agtaagtttt	ggagtatttg	aaataagttg	tgttgggtat	tttattttat	ggagagtgtg	1140
tagtgattga	tgtttattta	tagtgattag	agatgtatgt	tttgatagta	gtataaattt	1200
agtaggtgtg	aataaatggg	aaagagaaat	tgggtaaata	agtattatgg	tttttttagtt	1260
gagaaagtgg	gggtttttaa	aaggggtttt	tgttgataga	aagggatggt	taattattga	1320
aattgtagag	gggtgtgggt	tgggtgttga	gtgtgtagat	tatatattatg	gtgggtgattg	1380
ttttgtgttt	gggtgtgttt	gtataggtta	tgggtgtttg	gattatgttt	tttaggaata	1440
tttttagtat	ttttgtgagtt	tttttgtaga	tgaggttgga	gatgtgtttt	atgttgttgt	1500
gggtgagtaag	gtgttggatg	gttgggtttg	tgatgttttg	gatattgttg	tgtagtattt	1560
tatggtgggt	tttagtgttg	tttttgttaa	gatttttttt	gtttttgttg	tgggttagata	1620
tgatgagtaa	gaggagtttt	atttaattgt	ttgtgaggat	tttggtttga	ggtagtgttt	1680
ttatatgata	gttgggtggat	tgaattgaga	atttgaaaga	agttgggtgg	aagttttgtt	1740
ttgggtgggg	aggggaaatt	taaagggtta	aattgaaata	gggggaaaaa	aaaagtgagt	1800
tttttgggtt	tgtgttttga	attttgtaat	gtgtatagta	ttttgttatt	atgttatgag	1860
gttttaaaaa	attgtttttg	aatgtagaag	atataatatta	atattgtggg	aaatataaga	1920
aaggataaga	aattaagaaa	ttataatgtt	attttattat	ataggttagt	taattatgta	1980
ttttgtagag	tagttgtata	tatttttttt	agaaaatgta	tatagtgttg	tatatggagt	2040
tttgtaaatt	ttttatattg	attataattt	aattaaattt	tattaaagag	ataaaagtga	2100
tgttttgggt	tttatgtttt	ttaggaatta	tttaatagtt	taattagttt	tttagtaatt	2160
ttttaattgg	ttgtatttta	aaaataatgt	tttttatatt	taatataaat	gtattttttt	2220
tttatatttg	ggattaatat	tgaattttat	gattttatta	tattaaaaat	taaattttat	2280
tatattaata	tttaaaattg	tatttagaggt	tttatgattt	ggtattatgg	gtttttgtat	2340

tatttttttt	ttaaattttt	taatttggtt	tattaagggt	tttggataat	tttagagatt	2400
ttttgtgaag	tttgaataaa	atttttttga	gatttttgata	attgtattag	ttttaggatt	2460
taattggaat	agaattaaaa	tttttaaaat	aagtttttat	a		2501

<210> 16
 <211> 2501
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> chemically treated genomic DNA (Homo sapiens)

<400> 16

tataagaggt	tgttttaagg	attttaattt	tatttttaatt	aagtttttaa	gttaaatgtaa	60
ttattaaaa	tttgaagaga	ttttatttaa	atttttataaa	aggtttttaa	agttgttttag	120
aaatttttgg	gaaatagatt	aggaaatttg	gaaagggaat	aatgtggaga	ttttagtat	180
taaattatga	gatttttaaat	ataattttta	atattaatgt	aataaaattt	aaatttttgg	240
gtaataaaat	tataaatttt	aatatttggt	tttaagtata	agaaaaagta	tatttatggt	300
gaatgtggaa	aatattattt	ttaaaatata	gttgattaaa	aaattgttgg	ggaattgatt	360
ataattattg	ataattttta	agaaatatag	atattaaaa	attattttta	ttttttta	420
agaaattggt	taaattataa	ttaataataa	gaggttataa	aatttttatat	ataatattgt	480
atataatttt	ttggaaaaat	atgtgttaatt	gttttgtaaa	atataatgatt	aattagtttg	540
ttgtatggga	taatatgtta	gtttttta	tttttgtttt	tttttgtatt	ttttatagta	600
ttgatgtata	ttttttgtgt	ttaaaagtaa	tttttttaag	ttttataatg	tggtataaaa	660
atattatgta	tggtataaaa	tttagaatat	ggaaataaga	agtttgtttt	tttttttttt	720
ttatttttgg	ttgggttttt	agattttttt	ttttttattg	gggtgggatt	ttttgttgat	780
tttttttagg	tttttagttt	ggtttggtta	ttgtgtgata	aaggtgttgt	tttaggttag	840
agttttttta	aagtgttggg	tgagattttt	tttgtttgtt	atgtttgggt	gtggtaaaag	900
tggaagggtt	tttggtaaa	gtggtgttaa	gtgttattgt	aaagtattgt	gtgataatat	960
ttagggtatt	attaagtgtg	ttatttggtg	ttttgtttgt	tggtgtggtg	tgaagtgtat	1020
ttttggtttt	atttatgagg	agatttgtgg	gggtgtgaa	gtgtttttgg	agaatgtgat	1080
ttgggatgtt	gtgatttata	tagagtattg	taagtgtaa	atggttattg	ttatggatgt	1140
ggtttatgtg	tttaagtgtt	agggttgtat	tttttatggt	tttgggtgtt	gagtgttttt	1200
ttttattaat	aaaagggttt	tttttaggtt	tttttttttt	tagttgagga	gttggtgatg	1260
ttgtttgttt	agtttttttt	tattatttgt	ttgtgtttgt	tgagtttgtg	ttgttattgg	1320
agtatgtgtt	tttagtttgt	gtaagtaggt	attagttatt	aattgttttt	tagtaataaa	1380
aatatttta	ataatttgtt	ttagggtgtt	tagagtttat	tgatatgggt	tggtgtgtag	1440
attgtagtgt	agtgttatta	tggtttattg	tagttttttg	attagttgga	attttgttgt	1500
tattattatg	taaggttaat	tattttttta	aagattgggt	ttttgtgtgt	tttttaggtt	1560
agttttgata	ttttgggttt	aatttttttt	tttatgttta	atgttggtat	tatagtagtg	1620
agttattatg	tttgggttat	atattgttga	ggtttttagg	ttagttatat	tttaaggggt	1680
atttttgtag	tgtagtggg	aggaaagtta	agtagttata	ggtttttgg	gttggtgaat	1740
taattgttga	atatagtaa	aattttattt	gtaataattt	tttaattagt	gattgtaaat	1800
aagtttaggt	tattgggttt	ggtttaattt	ttgagaaagg	tgagaatttg	aatttttgag	1860
tagaaatatg	tagtttatta	gtataataaa	gtgattataa	taaagattaa	tttatttttt	1920
gttatttta	gtgggttag	ttttatttgt	gagaattttt	tagtttttga	gtgttttttt	1980
atttataatt	gaagtgtata	agatgggtat	aaaagtgaga	tttttagtaa	gtaaatttaa	2040
aggtttgaag	gattagtgt	taatgggaag	tgtaataaag	ttatatttga	ggttatttaa	2100
tttgagattt	tgatttaatt	ggtttaagga	tattattatt	ttgtgggtta	gatttgaaaa	2160
ataataattg	ttttaaggaa	ggaattttgt	ggtattttaa	taaaaaattt	ttttttttga	2220
tttttatatg	ttgttattta	gataaatttt	tggtttatat	ggttggattt	ttttttgtta	2280
gttaaatttt	gtgttatttt	gtgaaaaaga	attgagatta	ataaaatggt	agtgagaata	2340
taggggaaag	taaaaattat	agtaagttta	tttagttatg	ttgtagtatt	ttttataagg	2400
ataggggaag	attttgtttt	ttatttatta	tttaagtgt	gagtataaat	tttaaaaaata	2460
tatgatatta	taggaataat	gaagatgttt	ttatttgtaa	a		2501

<210> 17
 <211> 2251
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> chemically treated genomic DNA (Homo sapiens)

<400> 17

ggattttgag	ttgttttttag	gtagtttagga	gtgggttttag	tggttagttgt	aaatttttagt	60
agttgtttgt	gttgttttgt	gttgggggtg	agggtagtta	gagttgggga	tttaagggttt	120
gtgttatttg	tgtgtatagt	tttatatttg	aatgtttgtt	ttttgtagat	gagattgggtg	180
ggattgttaa	agttgggatt	tgtttttgaa	ggaaaaaaa	tagtgagtaa	gaaatttagt	240

attatttttt	attgatttat	tttgttgtat	tttttgtttt	ttaagttttt	gaaagttggg	300
aattttgatt	ttgggtgtta	aaaattgata	gttattgaga	ttgggtttga	gaagttgaag	360
atttggtagt	tttttagatt	agtaggataa	gggtgaaagta	gggttgagg	gggttttagga	420
tatttgaggg	ttgatttttg	gggtttgtga	gggtgttatt	gttgttgttg	ttataggtga	480
gatgggtgtg	gggtgatgtt	gggtttaatg	ggtagagaat	gtagggatgt	gggttttgtt	540
gaagagagtt	aagaaggga	gagtggtttt	tttaaattgt	ttttgtaatt	tggttttagt	600
gagtatttta	ttgatttaga	atttattgag	aatagtatta	gtgagttatt	ttttttttga	660
gatgggtttt	atttattttg	gtaatggagt	gagttggatt	gtggggagga	agaggaatgg	720
gaaaattagt	ttataaatat	taatgttagt	aagagtgtgt	tggttgtagg	atgtattgtg	780
agtttggaga	ttttgggtgt	tgtagtgtgt	aagtggttat	aatttagaaa	gtaggattga	840
gttgtttttt	ttgttttatt	agtgtattgt	tttttgggtg	tgggtttaat	attttataag	900
tggtaatttt	tgtttttggt	agttttgttt	tttttttatt	atttttagat	ttagttttgt	960
attttaagg	tgtgtattgt	tagttattat	tatgtttatt	tttgggggta	atttgtttgt	1020
ttttttgagt	tttgattgg	tgaatagttt	agtgtattat	ttgggtgtga	tgtttttttt	1080
tgggggtgg	ggtaatttgg	tgggttattgt	gggtgttgtgt	aagttgtgtg	aggagtagaa	1140
ggagatgatt	ttttatatgt	tgggtatgtg	gttgggtgtt	attgatttgt	tgggtatttt	1200
gttgggtgag	ttgggtgatta	ttgttatgta	tatgaagggt	taatgggttg	gggggttagtt	1260
gttgtgtgag	tattagtattt	ttattttgtt	tttttttagt	ttgtttgggt	ttagtattat	1320
ttgtgttatg	agtgttgagt	gttatttgg	tattaattat	gtttattttt	atagttatta	1380
tgtggataag	tgattgggtg	gttttatgtt	ttttgtagt	tatgtgttta	atgtgttttt	1440
ttgtgtgttg	tttaatatgg	gttttggtag	tttgtgggtg	tagtatttag	atatttgggtg	1500
ttttattgat	tggattatta	atgtgatgg	gtatgttgtt	tatttttata	tgtagtggg	1560
ttttagtttt	ttttttattt	ttgttattgt	tttttgaat	gtgtttgtgt	gtgggtgtgt	1620
gttttgtatg	tattgttagt	ttatgtgttg	tattttgttg	gggtattgag	agtattatgt	1680
gggtgtgggt	gttttgggtg	ttttttgggg	ttattttgtt	gttttttag	tttgttgtg	1740
tttttagtgat	ttttgggtgt	gttggagttt	ttgttgtatt	gtgggtgttg	agatttagat	1800
ggttattttta	ttttattgtta	tttttttgg	gggtgtttatt	tgttttattt	tgtttgtgtg	1860
gagtgatggg	gggtgggggt	ttatttgggt	ttttttttgt	atttattttt	tggttttatt	1920
ttttgttttt	tgtttttttt	ttgagttttt	ggtagtgaat	gtgtttgttt	taggttgggg	1980
ttgggatttt	tataattgtt	tttagagtag	gtttaatttt	ttttgaagtt	ttaattttaa	2040
tgagatttag	taggtgtttt	gtttttatat	tttttagttt	atgtttttgg	aagtttgggt	2100
tttttttttt	attgagatag	tttttatttt	ttgttgtttg	atattgggtg	agttttttta	2160
gaaaattttt	tggttttttt	gttagatgtg	gaggggagtt	tggtgtagtg	tgatttagtt	2220
tatttttttt	tattgtgaat	tgtgaattgt	a			2251

<210> 18

<211> 2251

<212> DNA

<213> Artificial Sequence

<220>

<223> chemically treated genomic DNA (Homo sapiens)

<400> 18

tgtagtttat	agtttatagt	atggaggaat	gggttaagtt	atattatagt	aggttttttt	60
ttatgtttaa	tagaggggg	gggggtttt	tttgaagat	ttggttagtg	ttaggtagta	120
aggggtaagg	gttgttttg	tggagaaaga	aatttaggtt	tttgggaata	taaattgggg	180
gatgtagggg	taaagtattt	gttaaatttt	gttaggggtg	ggattttaaa	gaggggtggg	240
tttgttttga	gaaatagtgt	gggaatttta	gttttgattt	aaaggtgata	tgttttattg	300
taaggattta	gagggaaagt	agggagtggt	gaatggatgt	gggaggtgga	tgtagaaaaa	360
aggtttagta	gggttttagt	tttgggtatt	tattatgagt	gggatggagt	agatgagtat	420
tattagggag	gtggtaaatga	gtaagatgat	tatttggatt	ttggtgtttg	tgatgtgggtg	480
gaagtttttg	tgggtgtgaa	agttgttgag	gtgtggtaag	gttggggagg	tagtgggggtg	540
gttttgggag	gtaattgagg	tgggtgtgg	tgtgtgggtg	tgtttgggtg	ttagttaggt	600
gtggtgtatg	aattgggtgg	gtatgtggag	tagtgtgtgt	tatataagta	tgttgttagg	660
gtatggtggg	agaatgagga	aggagtgtga	gttgtgtgat	atgtaggagt	agggtgtgtg	720
tgttgttatg	ttggtgtttt	agttgatgaa	gtattaggtg	tttgggtatt	gtagttgtga	780
gttattgaga	tttatgtttg	gtagtgtgta	aaagagtatg	ttggatgtat	agattgtaaa	840
gagtgtgagg	tttgttaatt	gtttgtttat	gtagtgggtg	tagaaatagg	tatgggtgat	900
ggttaggtag	tgtttgatat	ttatgggtga	gatgatgttg	aggttggata	ggttgaagaa	960
gagtagaatg	aagggtgtgt	atttgtatat	tgggtgtttt	ttgggttatt	ggttttttat	1020
gtatgtggtg	atgggtattg	ggtttattaa	taaagtgttt	aatagggttg	tgatagttag	1080
tttatatat	agtggtgtag	agggtgtttt	tttttgtttt	ttgtgtgatt	tgtatagtat	1140
tatgatgggt	attaggttgt	ttattatttt	gaagatgaat	attattgttg	ggatgggttat	1200
tgggttgttt	agttgggttg	ggtttaagga	gggtggatgaa	ttgatttttg	gagtggatat	1260
gatagtgggt	gggtgggtgt	agttttggag	tgtaagggtg	ggtttgggga	tggtagaaga	1320
gagataaagt	tgttgtgagt	ggaaattatt	atttgaaggt	tgtagatttt	gtgtttgaga	1380
aatgatatat	tgataagata	aggggagtaa	tttgttttta	tttttggat	tgtagttatt	1440
tattaattgt	ggtaattaaa	atttttaggt	ttgtgatatg	ttttgttaat	agtatatttt	1500
tgttgatatt	aatattttata	aattgatttt	tttatttttt	ttttttttta	taattttaatt	1560

tattttattg	ttaagatgaa	taagatttat	tttaagggaa	aagtagtttg	ttagtgttat	1620
ttttgataga	ttttgaatta	ataaaatgtt	tattgaaaat	aagttataaa	agtaatttgg	1680
agagtgtgtt	tttttttttt	tgggtttttt	tgggtgagggt	tgtatttttg	tgttttttat	1740
ttgttgattt	taatgttagt	ttaatgttat	tttatttgta	gtggtagtag	tggtggtagt	1800
tttatgagtt	tttagggtta	gttttttagt	gttttggatt	tgtttttaat	ttgtttttat	1860
tttgttttgt	ttagtttggg	aattgtttaa	tttttggttt	tttaaagttg	gttttagtgg	1920
ttgttgattt	ttggatattt	aggttagagt	tgtagttttt	taaaaatttg	ggaaataaga	1980
ggtgtagtgg	gatgggttag	tgaagaatgg	tggtggattt	tttatttggt	attttttttt	2040
ttttaaagat	gagtttttag	ttttagtgtt	ttgttgggtt	tgtttggggg	aggatagtgt	2100
ttaggtgtga	ggttgtgtgt	gtaggtgggt	tggagttttg	gttttttagt	ttggttggtt	2160
tttgtttttag	ttatgggtag	tatgggtggg	tattggagtt	tggtggttgt	gttgaggttg	2220
tttttgggta	tttgggggtg	gttttgggggt	t			2251

<210> 19

<211> 2586

<212> DNA

<213> Artificial Sequence

<220>

<223> chemically treated genomic DNA (Homo sapiens)

<400> 19

tttgtgtggg	tgtgggagta	gttttgttgg	gtgtttgtag	ttgtgggagt	taagtttttt	60
tttttagtgt	aggataaaaa	gtttatgggt	tttgaataat	gtggggtaga	ggttttttta	120
agtaatgttt	aattgggtgt	ttttaattaa	ggaaagagag	gttttttagt	ttatggtaat	180
ttaagtaggg	tagtttttag	ttaaagggtat	tttagaataa	taagattatt	ttaagaaatg	240
gaatgtttta	ttggatattt	gaataggttt	tttgttattg	gaattgggtg	gtattgtatt	300
ttaattagta	tttttgtgtg	gagggaggtg	atttagttta	ggaaagttaa	ttatagaaag	360
agggtgattt	tgaagagatt	gttttagtgt	tattagaata	tatgtgatta	tattaaaagt	420
ttaggtggat	attttagtgt	agtttggatt	tggataaatt	aatattgttg	gtagaattga	480
agggaaataag	ttattttta	tttatttttt	gtatgtgtag	tggtgagtga	gttgggggtg	540
ggaggatagt	ggtttgttta	ttgttttttt	ttaaaatttg	agatttgaaa	atatggagggt	600
tttatttgtt	tttttagttt	ttgattgtta	ataaaaaaat	aaattttgtt	ggttatatatt	660
tttttttttt	ttaaaatagt	aatttttatg	tttgtattaa	gttttttaga	agtttatatt	720
atttgttttt	ggttagggag	ggaataatgt	taggaaaagt	tattgggtgt	tttttatatt	780
tggttttttt	agggtgtagg	atgtgtgggt	tggtgggttt	gtgatttttg	aatgtttttt	840
gttatttttt	tgtgtgaatt	tgaagagatt	gggaggtgtt	gagagtagag	tttaggggtg	900
gtgtattttg	tgggtgttag	tgggtgggtg	gtttgggtgt	ttaggttggg	ggattttagt	960
ttgttttttt	gtggagggga	aaatatgtag	ttggaggggt	tgtgttgtgt	gggttgtgtg	1020
ttgttttttt	ttgttttttt	tttgaagtgt	tttaagtgtt	taggtagggg	tgtgagagtt	1080
tttttagagt	gaagtttttg	aggttgattt	gtgggttttg	ttgttatggg	aatttgggtg	1140
gtttaaagaa	gttttttttt	tgggtatttg	gaatttttag	ttagtgtggg	gtattgggga	1200
agtgggtgtg	gggggttggg	ttgggggatt	ttagttggta	gttttggaga	gggtttattt	1260
ttgggggtgt	tgggtgaggt	tgggtatgatt	tttggtttta	aaaggaaaagt	ttttgttttt	1320
tgtttttggg	tgagaagtta	aagattttatt	ttgagagtgg	agagagaaat	gttattggta	1380
atgttttttt	tggaaaagtt	gagaggggtt	ttttggatat	attatttagt	gtttttaaat	1440
tagagaagta	gttttttttt	ggtgttttgg	tttagaagtt	gttattttatt	tagtttatgg	1500
tttgaaatta	gtatgggaag	tgttggggta	aggttttgtt	ggagattaga	ggtttgtttg	1560
ttgggaggag	tttttggggg	atggggattt	tatttttttg	tttgttttgg	tttttatattg	1620
ggatgttttt	gtaggagttt	agaaagatga	tttattatat	ggttttggga	tagagttagt	1680
tgttttaatt	tgagggaatt	ttgtgtgttt	ttttgaggtt	ttagtttttt	aaggatttgt	1740
tgtttgtttt	ttttttttta	gattgaaatt	ggggaagagt	gtgggtgttt	ttttgttttg	1800
atgagtttgt	tttttttaaat	gtttattttg	gttgtatttag	agtatttggg	aaattttgaa	1860
agggtgtttag	gttttatata	gtagtgtttt	tttatttagt	ttttgttttt	gggttttttt	1920
aagagagttt	ttattttatg	ttttggtttt	tttttgatgt	tgggtttttg	aggtaggtat	1980
ggagtttttt	tgaagtagt	tgttttattg	tgtttttttg	gtgtaaaagt	agagttttatt	2040
ttgttggggg	aaggggaggt	agaaaagatt	atagttggga	aagtgtgttt	tttgttttgt	2100
ttttaaaata	tgtttttaaga	ttgttattgt	gattgttagg	agagttatta	atgttttaggg	2160
gttataaagg	aatttttgaa	tttttgggtt	tttttaaat	tttaggtttt	taaaatttta	2220
gtgggggttt	tttgggggtg	ggatttaggt	tggatttgtt	gggaggattt	tgtttagtat	2280
tttttttata	atattttcat	aaggtaggtt	ttgttttttt	ttggagtttt	tttttttggg	2340
atgttttttaa	attttgggtta	atttattttt	ttgtgagttta	ttttaggggt	ttgtgggttg	2400
ggaagagatg	tgtttaatttt	gtgggtttgt	gtgtagtttt	tagttgtaaa	gtgttgtaag	2460
tgattttttt	gatgggtttt	tttgattgaa	gagtttggga	attaaagaga	aaaaaataa	2520
ttttattttt	aaaagaataa	gttattattg	tgggtgatatt	gtggtggagg	attttggtga	2580
tgggggt						2586

<210> 20

<211> 2586

<212> DNA

<213> Artificial Sequence

<220>

<223> chemically treated genomic DNA (Homo sapiens)

<400> 20

```
atattattgt taaagttttt tgttatagta ttgttgtagt gatgatttgt ttttttgaaa 60
ataaagttat tttttttttt tttgggtttt gagttttttg gttggaaagg gttgttaggg 120
gggttatttg tagtattttg tggttaagga ttgtgtgtag atttgtggag ttgatgtgtt 180
tttttttggg ttatagagtt ttaggatggg ttatagggga gtgagttagt tagagtttgg 240
gaatattttg agaaaagagg ttttagagaa ggtaggagtt tgtttttgtg aaatattaat 300
aaagggatgt taggtgaggt ttttttagtg gtgttagttt gaattttagt ttttaagagat 360
ttttattagg gtttttaggaa tttgggggtt tgggaagggt tgagggttta gaaatttttt 420
tatagttttt agatgttgat agttttttta ataattgtga tgatagtttt gaggtatggt 480
ttaggaataa ggtgaaaagt gtattttttt aattgtgatt ttttttattt tttttttttt 540
taataaagta aatttttaatt ttatattaaa ggggtataga taggtaattg ttttttagagg 600
gattttgtgt ttattttggg gatttgatat tgaagaaaga ttgaggtatg aggtagagat 660
ttttttgaaa aaattttaaag atagaggttg agtagggaga tgttgttgtg tgagggttgg 720
gtatttttta gagtttttta gatgttttgg tgtagttaga gtaggtgttt ggggaggtga 780
atttattggg gtaaaagaag gtttatatat ttttttagtt ttgggttagg gaaagaagaa 840
tggatgggtg tgttttggaa agttaggggt ttagagaggt gtataaagtt tttttaaagt 900
tgggtgtgtt gttttgtttt gggattatgt agtggattgt ttttttgggt ttttatggag 960
gtgttttagg tgggaattag agtaagtagt agaattgggt ttttattttt taggggtttt 1020
ttttgatagg taggttttta gtttttgatg aagttttgtt ttggtgtttt ttatgttgat 1080
tttgaattat ggggttgagt agtggtgatt tttgagttta ggtattaaag aaaaattatt 1140
tttttgggtt ggggggtatta gtagtgtgtt ttagaagatt ttttttgaat tttttaaaga 1200
aaatgttatt aataatattt ttttttttgt ttttaaaata agtttttggg tttttgtgtt 1260
agaataagaa gtagaaattt tttttttgga gtttaagaatt gtgttggttt tatttagtga 1320
ttttaagggt aggttttttt tggagttgtt gggtgaggtt ttttaattta gttttttagt 1380
gttatttttt tgatgtttta tattaatttg gaatttttag tgtttggaag aaagggtttt 1440
ttggattaat tgggttttta tagtagttag atttataggt tagtttttgg ggttttagtt 1500
ttaggggatt tttatagttt tatttagagg tttggggaga ttttaggatg attgatgggg 1560
taatggatta taatttgtat ggtgtatgtt ttttagttat gtattttttt ttttgtggta 1620
gggtgattat aggttttttt gtttgagtgt ttgggtgtgt tgtttattta gtattgtaga 1680
gtgtattagt tttgaatttt gtttttaata ttttttagtt tttttaagtt tgtgtaaggg 1740
gatggttaga agtgttttgg gattataggt ttgttggttt gtataattttg tattttggaa 1800
ggggtgagga tggaaagagta ttgggtgatt ttttagtat tgtttttttt ttgggtttaga 1860
gtaaatgaga taaatttttt aaggggttaa tataagttat aggggtgtta ttttggaaatg 1920
aggagaaaat gtagtttagt ggatttgggt ttttgttggg aattaagagt tgggaaaatg 1980
aatgggattt ttatattttt agatttttag ttttaaaagg gggtaaataa tgaattgttg 2040
tttttttatt ttaattttat ttagtattat gtgtataggg gatgggggtg gggtaatttg 2100
tttttttttag ttttgttagt aatgttgaat tgtttaaatt tgagttggtt gtgggtgttt 2160
gtttgggttt ttggtgtggt tatatgtatt ttaatagtgt taagataatt tttttggagg 2220
ttattttttt ttgtagttga tttttttaga ttgggttgtt tttttttata taagagtatt 2280
ggttgaagta tagtatatat taattttaat gatagagaat ttggttgggt gtttagtgag 2340
atattttatt ttttagaatg attttattat tttaaagtat ttttagtttg aagttgtttt 2400
gtttgggttg ttatagagtt ggaagttttt ttttttttaa ttagaagtgg ttaattagat 2460
gttgttttga aaaatttttg ttttgtattg ttttaagagt ataaattttt atgtttgtat 2520
ttggggaggg ggtttgattt ttgtggttgt gagtgtttag tggggttatt tttgtgtttg 2580
tgtggg 2586
```

<210> 21

<211> 20

<212> DNA

<213> Artificial Sequence

<220>

<223> chemically treated genomic DNA (Homo sapiens)

<400> 21

gggattattt ttataagggt

20

<210> 22

<211> 17

<212> DNA

<213> Artificial Sequence

<220>

<223> chemically treated genomic DNA (Homo sapiens)

<400> 22
 ctctaaaccc catcccc 17
 <210> 23
 <211> 24
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> chemically treated genomic DNA (Homo sapiens)
 <400> 23
 cccatcccca aaaacacaaa ccac 24
 <210> 24
 <211> 19
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> chemically treated genomic DNA (Homo sapiens)
 <400> 24
 cgtcgtcgta gttttcggt 19
 <210> 25
 <211> 18
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> chemically treated genomic DNA (Homo sapiens)
 <400> 25
 tagtgagtac ggcgcggt 18
 <210> 26
 <211> 19
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> chemically treated genomic DNA (Homo sapiens)
 <400> 26
 accgaaaata cgcttcacg 19
 <210> 27
 <211> 22
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> chemically treated genomic DNA (Homo sapiens)
 <400> 27
 gcgttatcgt aaagtattgc gc 22
 <210> 28
 <211> 19
 <212> DNA
 <213> Artificial Sequence
 <220>

<223> chemically treated genomic DNA (Homo sapiens)

<400> 28

cgcgacgaac aaaacgccg

19

<210> 29

<211> 18

<212> DNA

<213> Artificial Sequence

<220>

<223> chemically treated genomic DNA (Homo sapiens)

<400> 29

cgcgctactc cgcataca

18

<210> 30

<211> 19

<212> DNA

<213> Artificial Sequence

<220>

<223> chemically treated genomic DNA (Homo sapiens)

<400> 30

gaggtaatcg aggcggtcg

19

<210> 31

<211> 21

<212> DNA

<213> Artificial Sequence

<220>

<223> chemically treated genomic DNA (Homo sapiens)

<400> 31

cgccaattca tacgccgcac c

21

<210> 32

<211> 19

<212> DNA

<213> Artificial Sequence

<220>

<223> chemically treated genomic DNA (Homo sapiens)

<400> 32

ttgtggttcg ggaagagac

19

<210> 33

<211> 18

<212> DNA

<213> Artificial Sequence

<220>

<223> chemically treated genomic DNA (Homo sapiens)

<400> 33

tcccgaactc ttcgatcg

18

<210> 34

<211> 20

<212> DNA

<213> Artificial Sequence

<220>

<223> chemically treated genomic DNA (Homo sapiens)

<400> 34

aactacgcgc aaacccgcga

20

